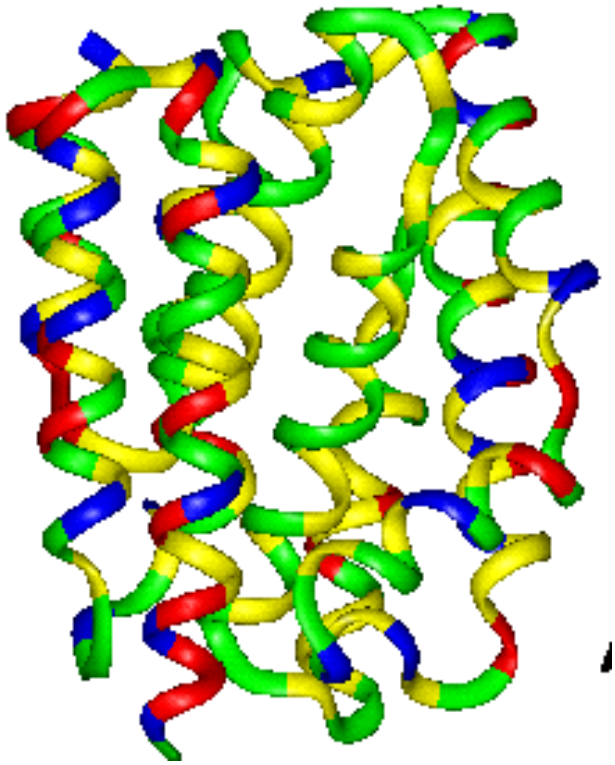
Medical Biochemistry
Lecture Notes
2002-2003



Book I-A
Proteins: Structures & Function

The USUHS Medical Biochemistry Course

This course is designed to provide you with the biochemistry background for basic science and clinical courses at USUHS, and to prepare you for the National Board Exams at the end of the second year. It will also give you the background to keep abreast of advances in molecular medicine.

The course is divided into four major areas:

1. Proteins and enzymes

2. Medical genetics and Nucleic Acids

3. Metabolism:
   A. Carbohydrate metabolism
   B. Lipid metabolism
   C. Amino acid and nitrogen metabolism

4. Nutrition

The syllabus for the entire course is provided at the end of this introductory section.

Lectures-Medical Biochemistry is a lecture-based course. Lectures are where the concepts, essential facts, and biomedical relevance of biochemistry are presented and explained.

All instructors provide handout material to accompany their lectures. Though some of these handouts may appear to contain all you need to know, instructors regard them as an adjunct to lectures, not a substitute. You are responsible for what is presented in class.

Clinical Correlations-These are lectures devoted to specific diseases and disorders. Most of the clinical correlations are given by outside speakers, who are both physicians and researchers.

Recitations-These sessions, led by a single instructor, are devoted to working through problems and exercises that have been assigned as homework. Recitations are held in the lecture hall.

Textbook Assignments-Our main textbook is "Principles of Biochemistry " by Lehninger, Nelson, and Cox. It is an excellent text of basic biochemistry, with a strong emphasis on human biochemistry. A good, user-friendly text of medical biochemistry is "Basic Medical Biochemistry" by Marks, Marks, and Smith, which is on reserve in the LRC. Each text has its strengths and weaknesses. Lehninger is better for the basic concepts, whereas Marks does a better job integrating intermediary metabolism with medical physiology, and providing a clinical context for the material. The clinical correlations in Marks are excellent. Most instructors regard the texts as reference sources, although in a few cases sections may be assigned as required reading. Some students prefer a basic text (such as Lehninger or the one they used in undergraduate courses) whereas others prefer a more clinically oriented text. See what works for you.

Exams-Five hourly exams are given, covering the following blocks: 1) Proteins and Enzymes, 2) Genetics and Nucleic Acid Metabolism, 3) Carbohydrate Metabolism, 4) Lipid Metabolism, and 5) Amino Acid and Nitrogen Metabolism. Each exam consists of about 40-45 multiple choice and matching questions. There are usually 1-3 questions per lecture, and clinical correlation. A final comprehensive exam, obtained from the National Board of Medical Examiners, is given at the end of the course.

BIOCHEMISTRY GRADING POLICY

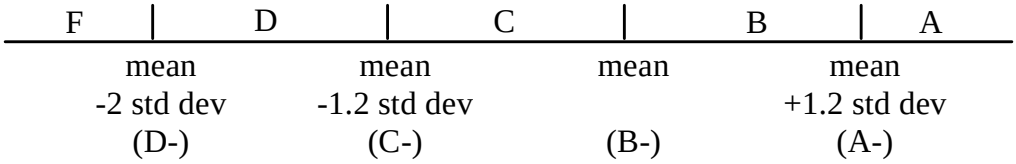
You will receive a single letter grade at the end of the course based upon your performance on the hourly exams and the final. The hourly exams count for 75% of your grade, and the final exam counts for 25%. The weight assigned to each hourly is roughly proportional to the amount of material covered, as follows:

	Maximum Points
Exam I	14
Exam II	18
Exam III	15
Exam IV	13
Exam V	15
Final Exam	25
	100 = Maximum Total

All exams consist of multiple-choice and matching National Board-style questions. Performance on each hourly exam will be recorded as the percentage of questions answered correctly. (There is no penalty for guessing.)

The final exam is a National Board of Medical Examiners "shelf" exam--an exam that was once given to second-year students as part of the USLME Step I. Your completed exams are sent back to the NBME for grading, and the scores are returned to us in the form of numbers (300-800) that reflect your percentile rank compared with other medical students across the country who took the exam under the same circumstances, i.e., as a comprehensive final for a biochemistry course in the same year. To convert National Board scores to our grading system, we set a score of 700 (approximately 99th percentile nationally) as 100% and apply the following formula: [(your score/700) x 100].

The final letter grade for the course will be computed according to the following statistical analysis:



However, independent of the class mean and standard deviation, the scores shown below guarantee you the corresponding grades:

Total Points Achieved	Letter Grade
86-100	A
76-85	B
66-75	C
56-65	D
55 or less	F

The class average is usually close to 76.

During an exam, typographical errors or obvious mistakes such as missing pages may be brought to the attention of a proctor; any corrections will be announced to the whole class. Proctors cannot clarify questions for individual students. After an exam, you may appeal to have a question re-graded or thrown out if you feel that (1) the posted answer is wrong, (2) there is no correct answer to choose from, (3) there is more than one correct answer, or (4) the question is unclear or ambiguous. Challenges must be made **in writing** to the Course Director or Biochemistry department office (B4058) within 48 hours after the answers have been posted. The Course Director will consult with the exam writer(s) and decide which action to take. If a challenge is accepted as valid, all students who stand to gain will receive credit, even if they did not submit a challenge themselves.

If you arrive at school late for an exam, report to your proctor **immediately**.

* * * * * * *

Your feedback is extremely valuable in designing next year's course. (You are the beneficiaries of suggestions made by previous classes.) Therefore, the faculty requires you to complete a course evaluation the end of the semester. Since this is a required part of the medical biochemistry course, your grade will stand as an "I" until the assignment is completed.

For routine administrative assistance, contact:
Biochemistry Office, B4058
295-3550
Melanie Joy, mjoy@usuhs.mil or Melissa Horn mehorn@usuhs.mil.

All other problems, contact:
Mark Roseman
Course Director, B3030
295-3570
mroseman@usuhs.mil.

BOOK I-A

PROTEINS: STRUCTURE AND FUNCTION

Date	Time	Topic	Instructor
Mon 8/26 Lec Rm E	0930-1020	(a) Introduction to Medical Biochemistry Course (b) The Molecules of Life	Roseman
Tues 8/27	1300-1350	Amino Acids: Structures & Properties	Roseman
Lec Rm E	1400-1450	pH and pKa; Acids, Bases, and Buffers	Roseman
Wed 8/28	1300-1350	Ionization of Amino Acids and Peptides	Roseman
Lec Rm E	1400-1450	Noncovalent Forces Between Molecules & Properties of Water	Roseman
Thurs 8/29 Lec Rm E	0730-0820	Clinical Correlation: Acidosis & Alkalosis	Cloonan
Fri 8/30	0930-1020	Protein Structure I	Dey
Lec Rm E	1030-1120	Protein Structure II	Dey

BOOK I-B

PROTEINS: STRUCTURE AND FUNCTION

Date	**Time**	**Topic**	**Instructor**
Mon 9/2		LABOR DAY HOLIDAY	
Tues 9/3	1400-1450	Binding Equilibria	Roseman
Lec Rm E	1500-1550	Ligand Binding to Proteins	Roseman
Wed 9/4	0930-1020	Thermodynamics	Roseman
Lec Rm E	1030-1120	Myoglobin and Hemoglobin I	Xiang
	1300-1350	Myoglobin and Hemoglobin II	Xiang
Thurs 9/5	0930-1020	Enzymes I	Dey
Lec Rm E	1030-1120	Enzymes II	Dey
Fri 9/6	0930-1020	Enzymes III	Dey
Lec Rm E	1030-1120	Enzymes IV	Dey
Mon 9/9	1030-1120	Muscle I	Xiang
Lec Rm E	1300-1350	Muscle II	Xiang
Tues 9/10	1000-1200	Separation and Purification of Proteins	Horowitz
Lec Rm E			
Wed 9/11	0730-0820	Review	
	0830-0920	Review	
Thurs 9/12	0730-0820	EXAM I	
Lec Rm E			

BOOK II-A

NUCLEIC ACIDS AND GENETICS

Date	**Time**	**Topic**	**Instructor**
Fri 9/13	1300-1350	Recitation: DNA Structure	Dunn
Lec Rm E	1400-1450	Recitation: DNA Structure	Dunn
Mon 9/16	1030-1120	Principles of Replication	Dunn
Lec Rm E	1300-1350	DNA Replication	Dunn
Tues 9/17	1400-1450	DNA Repair I	Dunn
Lec Rm E	1500-1550	DNA Repair II	Dunn
Wed 9/18	1030-1120	DNA Structure	Dunn
Lec Rm E			
Thurs 9/19	0930-1020	RNA Synthesis	Dunn
Lec Rm E	1030-1120	RNA Processing	Dunn
Fri 9/20	1300-1350	Genetic Code	Dunn
Lec Rm E			
Mon 9/23	0730-0820	Clinical Correlation: DNA Repair	Kraemer
Lec Rm E	1300-1350	Protein Synthesis I	Dunn
	1400-1450	Protein Synthesis II	Dunn
Tues 9/24	1400-1450	Gene Regulation I	Dunn
Lec Rm E	1500-1550	Gene Regulation II	Dunn
Wed 9/25	0730-0820	Gene Regulation III	Dunn
Lec Rm E			

BOOK II-B

NUCLEIC ACIDS AND GENETICS

Date	Time	Topic	Instructor
Thurs 9/26	1300-1350	Molecular Biology I	D'Arpa
Lec Rm E	1400-1450	Molecular Biology II	D'Arpa
Fri 9/27	1300-1350	Molecular Biology III	D'Arpa
Lec Rm E	1400-1450	Molecular Biology IV	D'Arpa
Mon 9/30	0730-0820	Medical Genetics I	Haynes
Lec Rm E	0830-0920	Medical Genetics II	Haynes
Tues 10/1	0730-0820	Medical Genetics III	Haynes
Lec Rm E	1400-1450	Population Genetics	Haynes
Wed 10/2	1300-1350	Recitation: Hemoglobinopathies	Haynes
Lec Rm E	1400-1450	Recitation: Hemoglobinopathies	Haynes
Thurs 10/3	1300-1350	Neurogenetics	Fischbeck
Lec Rm E	1400-1450	Molecular Medicine I	Haynes
Fri 10/4	1300-1350	Recitation: Genetics of Human Disease	Haynes
Lec Rm E	1400-1450	Recitation: Genetics of Human Disease	Haynes
Mon 10/7	0730-0820	Molecuar Medicine II	Haynes
Tues 10/8	1400-1450	Review	
Lec Rm E	1500-1550	Review	
Wed 10/9	0730-0820	EXAM II	
Lec Rm E			

BOOK III-A

CARBOHYDRATE AND LIPID METABOLISM

Date	Time	Topic	Instructor
Wed 10/9	1300-1350	Carbohydrate Structure & Properties	Roseman
Lec Rm E	1400-1450	Lipid Structure and Physical Properties I	Roseman
Thur 10/10	0730-0820	Lipid Structure and Physical Properties II	Roseman
Lec Rm E			
Fri 10/11	0730-0820	Glycoproteins	Grahame
Lec Rm E			
Mon 10/14		COLUMBUS DAY	
Tue 10/15	1400-1450	Introduction to Membranes	Grahame
Lec Rm E	1500-1550	Membrane Transport	Grahame
Wed 10/16	0930-1020	Protein Targeting	TerBush
Lec Rm E			
Thur 10/17	1300-1350	Bioenergetics: Coupled Reactions	Grahame
Lec Rm E	1400-1450	Bioenergetics: Oxidation Reduction	Grahame
Mon 10/21	0730-0820	Recitation: Bioenergetics I	Grahame
Lec Rm E	0830-0920	Recitation: Bioenergetics II	Grahame

BOOK III-B

CARBOHYDRATE AND LIPID METABOLISM

Date	**Time**	**Topic**	**Instructor**
Tues 10/22 Lec Rm E	1030-1120	Nutrition: Macronutrients and Digestion	Roseman
Wed 10/23 Lec Rm E	0730-0820	Glycolysis	Roseman
Thurs 10/24	1030-1120	Pyruvate Oxidation and TCA Cycle	Roseman
Lec Rm E	1300-1350	Mitochondrial Electron Transport	Grahame
	1400-1450	Oxidative Phosphorylation	Grahame
Fri 10/25 Lec Rm	0730-0820	Mitochondrial Membrane Transport	Roseman
Mon 10/28	0930-1020	Gluconeogenesis	Roseman
Lec Rm E	1030-1120	Glycogen Metabolism I	Roseman
Tues 10/29	1400-1450	Glycogen Metabolism II	Roseman
Lec Rm E	1500-1550	Hexose Monophosphate Pathway	Roseman
Wed 10/30 Lec Rm E	0730-0820	Regulation of Carbohydrate Metabolism Part I-Liver	Roseman
Thurs 10/31	1300-1350	Regulation of Carbohydrate Metabolism	Roseman
Lec Rm E		Part II-Muscle	
	1400-1450	Alternate Pathways of CHO Metabolism	Roseman
Fri 11/1 Lec Rm E	0730-0820	Review	
Mon 11/4 Lec Rm E	0730-0820	EXAM III	

BOOK III-C

CARBOHYDRATE AND LIPID METABOLISM

Date	Time	Topic	Instructor
Mon 11/4	1300-1350	Fatty Acids I	TerBush
Lec Rm E	1400-1450	Fatty Acids II	TerBush
Tues 11/5	1030-1120	Fatty Acids III	TerBush
Lec Rm E			
Wed 11/6	0730-0820	Lipoproteins I	TerBush
Lec Rm E			
Thurs 11/7	0930-1020	Lipoproteins II	TerBush
Lec Rm E	1030-1120	Lipoproteins III	TerBush
Fri 11/8	0930-1020	Cholesterol and Bile Acid Metabolism	TerBush
Lec Rm E	1030-1120	Steroid Hormone Synthesis	TerBush
Mon 11/11		VETERAN'S DAY	
Tues 11/12	0730-0820	Phospholipids and Sphingolipids	TerBush
Lec Rm E	1400-1450	Eicosanoid Synthesis	TerBush
Wed 11/13	0730-0820	Recitation: Glucose 6-Phosphatase Deficiency	Roseman
Lec Rm E	1300-1350	Clinical Correlation: Hyperlipoproteinemias	Remaley
Thurs 11/14	0730-0820	Lipid Storage Diseases	Dunn
LecRm E	1300-1350	Integration of CHO and Fat Metabolism I	Roseman
	1400-1450	Integration of CHO and Fat Metabolism II	Roseman
Fri 11/15	0730-0820	Amino Acid Transport	Roseman
Lec Rm E			
Mon 11/18	1300-1350	Excretion of Excess Nitrogen I	Roseman
Lec Rm E	1400-1450	Excretion of Excess Nitrogen II	Roseman
Tues 11/19	1300-1350	Review	
Lec Rm E			
Wed 11/20	0730-0820	EXAM IV	
Lec Rm E			

BOOK IV

NITROGEN METABOLISM AND INTEGRATION OF METABOLISM

Date	Time	Topic	Instructor
Wed 11/20 Lec Rm E	0830-0930	Amino Acid: Synthesis & Degradation	Roseman
Thurs 11/21 Lec Rm E	0730-0820	Specialized Products From Amino Acid Metabolism	Roseman
Fri 11/22	1300-1350	Purine and Pyrimidine Metabolism	Grahame
Lec Rm E	1400-1450	Disorders of Purine & Pyrimidine Metabolism	Grahame
Mon 11/25	0930-1020	Water Soluble Vitamins I	Horowitz
Lec Rm E	1030-1120	Water Soluble Vitamins II	Horowitz
Tues 11/26 Lec Rm E	0730-0820	Water Soluble Vitamins III	Horowitz
Wed 11/27 Lec Rm E	0730-0820	Nutrition: Fat Soluble Vitamins	Horowitz
Thurs 11/28		THANKSGIVING	
Fri 11/29		THANKSGIVING	

Book V

SPECIAL TOPICS

Date	Time	Topic	Instructor
Mon 12/2	1300-1350	Nutrition: Minerals	Horowitz
Lec Rm E	1400-1450	Clinical Correlation: Disorders of Amino Acid Met.	Francis
Tues 12/3	0900-0950	Intertissue Relationships in Amino Acid Metabolism	Roseman
Lec Rm E	1000-1050	Integration of Metabolism	Roseman
	1400-1450	Nutrition: Obesity	Schneider
Wed 12/4	1300-1350	Nutrition: Oxidative Damage	Horowitz
Lec Rm E	1400-1450	Red Cell Disorders	Laughlin
	1500-1550	Red Cell Disorders	Laughlin
Thurs 12/5	0930-1020	Cell Signaling I	D'Arpa
Lec Rm E	1030-1120	Cell Signaling II	D'Arpa
Fri 12/6	1300-1350	Heme Metabolism	Grahame
Lec Rm E	1400-1450	Disorders of Heme Metabolism	Grahame
Mon 12/9	0730-0820	Signal Transduction in Vision	Horowitz
Lec Rm E	1300-1350	Malnutrition	Poth
	1400-1550	Malnutrition	Poth
	1500-1550		
Tues 12/10	0730-0820	Collagen	Grahame
Lec Rm E	1300-1350	Clinical Correlation: Diabetes I	LeRoith
	1400-1450	Clinical Correlation: Diabetes II	LeRoith
Wed 12/11	0930-1020	Proteoglycans	Grahame
Lec Rm E	1030-1120	Clinical Correlation: Mucopolysaccharidosis	Conran
Thur 12/12	0730-0820	Review	
Lec Rm E			
Fri 12/13	0730-0820	EXAM V	
Lec Rm E			
Thurs 12/20	0830-1120	FINAL EXAM	

BOOK I-A

PROTEINS: STRUCTURE AND FUNCTION

Date	Time	Topic	Instructor	Chapter
Mon 8/26 Lec Rm E	0930-1020	(a) Introduction to Medical Biochemistry Course (b) The Molecules of Life	Roseman	1
Tues 8/27	1300-1350	Amino Acids: Structures & Properties	Roseman	2
Lec Rm E	1400-1450	pH and pKa; Acids, Bases, and Buffers	Roseman	3
Wed 8/28	1300-1350	Ionization of Amino Acids and Peptides	Roseman	4
Lec Rm E	1400-1450	Noncovalent Forces Between Molecules & Properties of Water	Roseman	5
Thurs 8/29 Lec Rm E	0730-0820	Clinical Correlation: Acidosis & Alkalosis	Cloonan	
Fri 8/30	0930-1020	Protein Structure I	Dey	6
Lec Rm E	1030-1120	Protein Structure II	Dey	6

Introduction to Medical Biochemistry:

The Molecules of Life

I. The major classes of complex biomolecules

A. Proteins–macromolecules (polymers) are chains of amino acids.

1. Structure of an amino acid

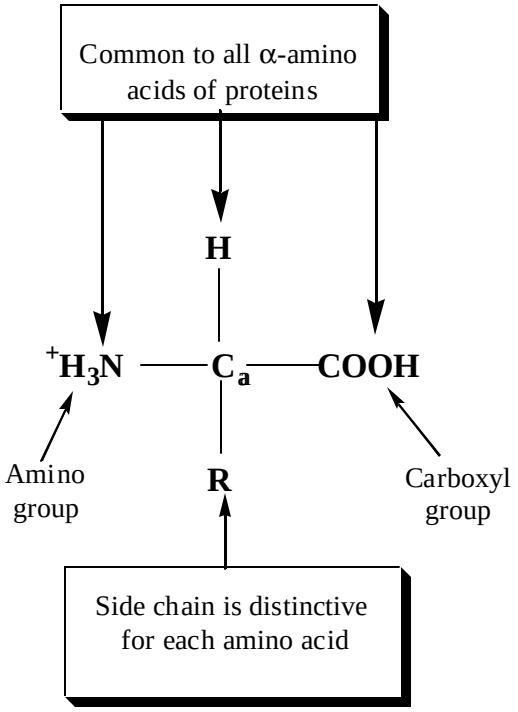
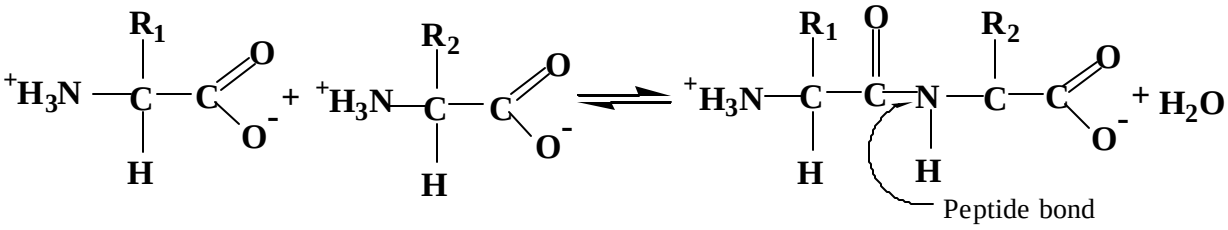


Figure 1: Structural features of amino acid (shown in their fully protonated form).

2. Linkage of amino acids via peptide (amide) bonds to form peptides and polypeptide chains.

Figure 2: Formation of a peptide bond.



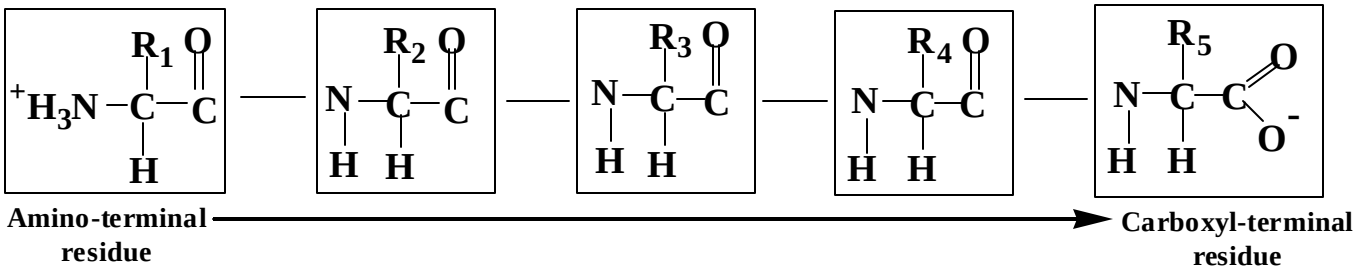
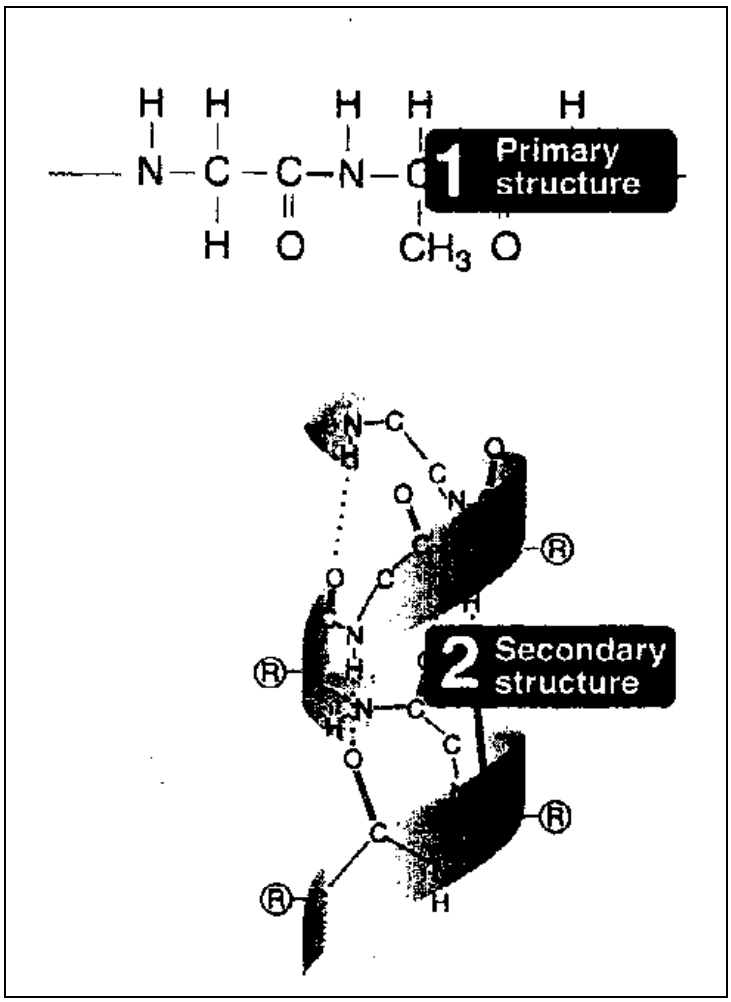


Figure 3: A pentapeptide. The constituent amino acid residues are outlined. The chain starts at the amino end.

3. Folding of chains into complex shapes.



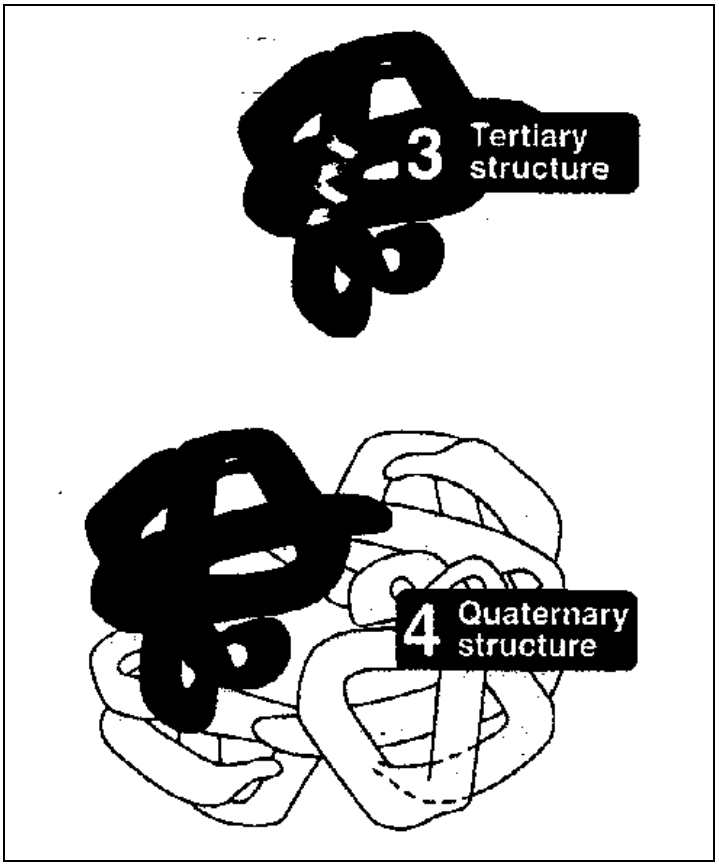
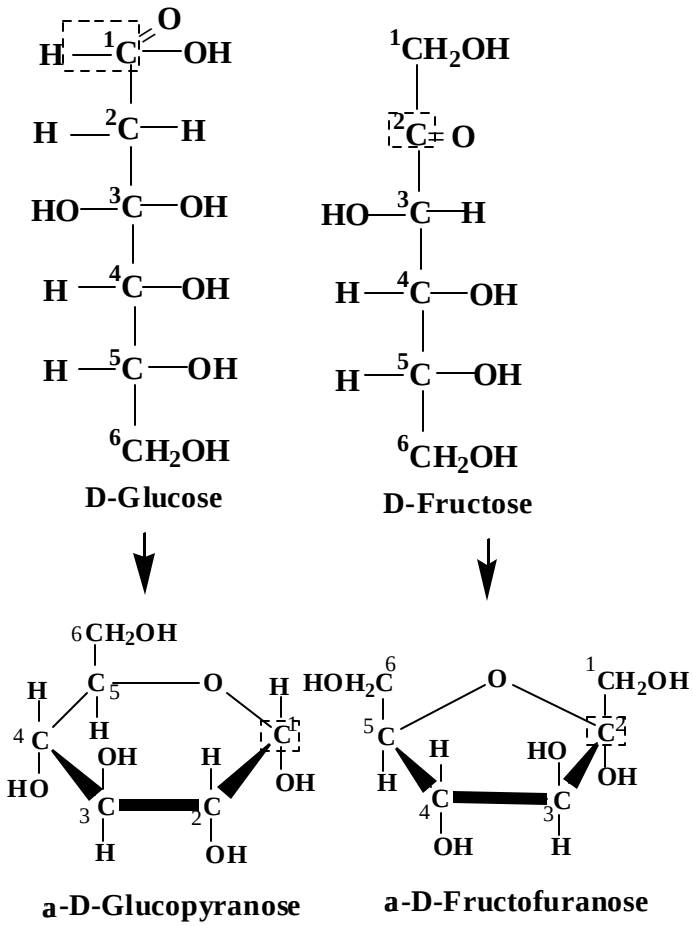


Figure 4: Four levels of protein structure.

B. Complex Carbohydrates–polymers of monosaccharides

1. Structures of typical monosaccharides



2. Linkage of monosaccharides via glycosidic (acetal) bonds to form oligosaccharides and polysaccharides.

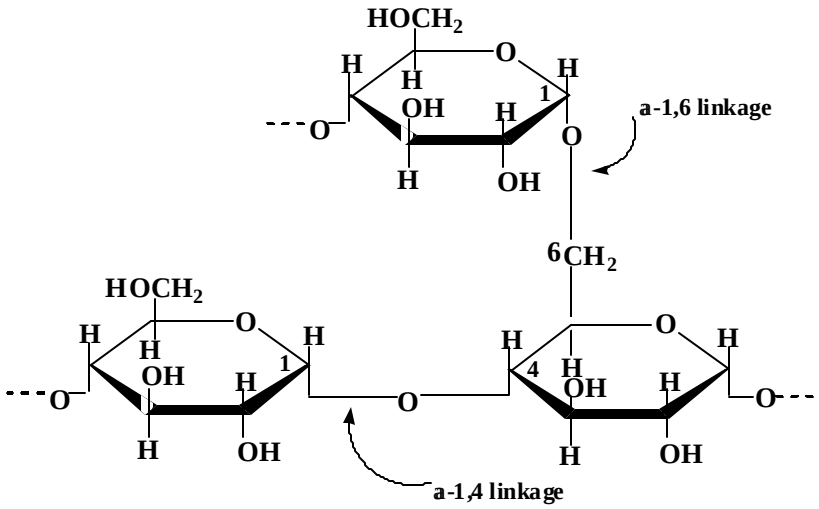


Figure 5: "-1,4 and "-1,6 linkages between glucose residues in starch and glycogen.

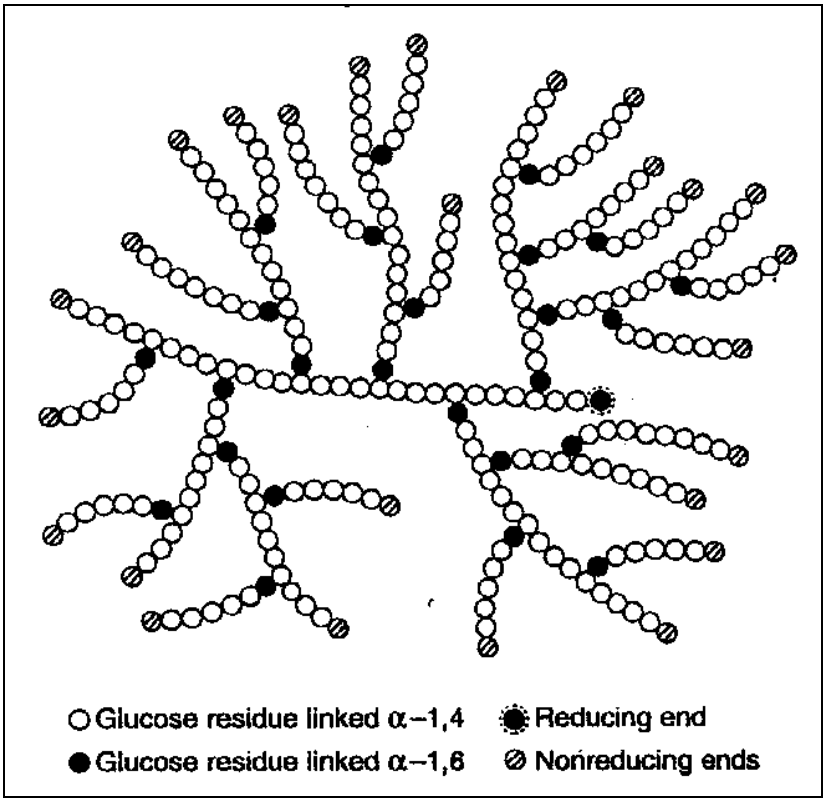
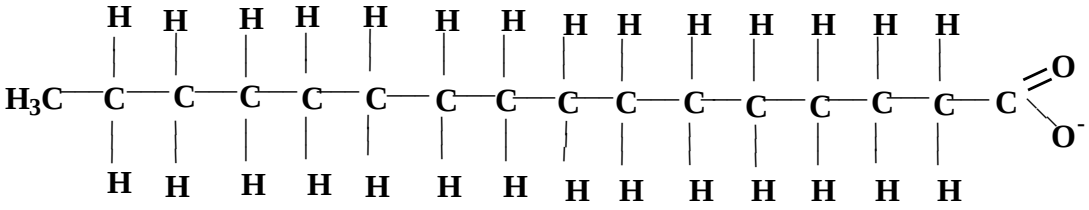
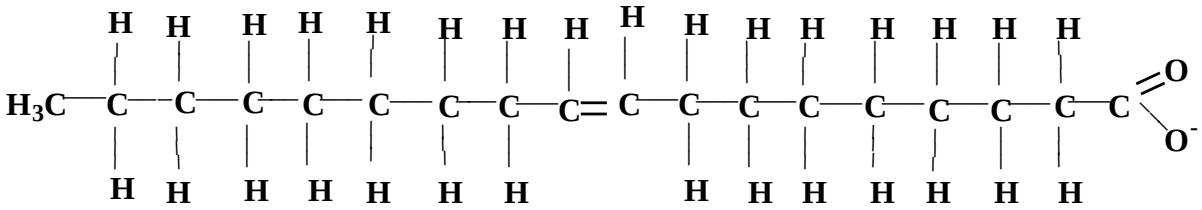


Figure 6: Structure of glycogen.

C. Lipids-structurally diverse hydrophobic/amphipathic molecules. Hydrophobic refers to a nonpolar, "water-hating" portion of a molecule whereas hydrophilic refers to a polar or charged "water loving" portion of a molecule. A molecule with separate hydrophobic and hydrophilic regions is called **amphipathic**. Examples are fatty acids shown below.



Palmitate
(Ionized form of palmitic acid)



Oleate
(Ionized form of oleic acid)

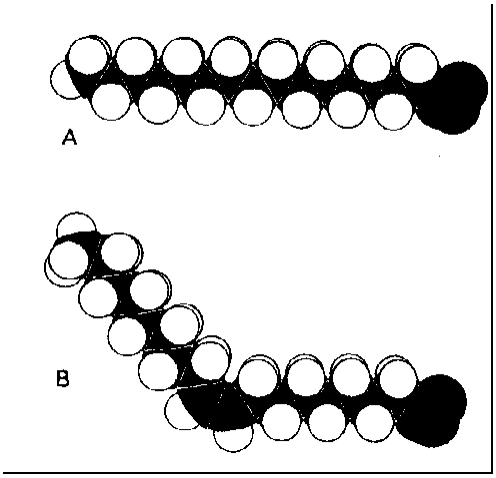


Figure 7: Space-filling models of (A) palmitate (C16, saturated) and (B) oleate (C18, unsaturated). The cis double bond in oleate produces a bend in the hydrocarbon chain.

The most important class of lipids is probably the phosphoglycerides, which form bilayer membranes. The simplest member of this family is phosphatidic acid, shown below, in which a polar phosphate group and two fatty acyl chains are connected to a glycerol backbone. Most phospholipids also have a water soluble alcohol attached to phosphate, such as in phosphatidylcholine (below). The linkage between the components are ester and phosphoester bonds.

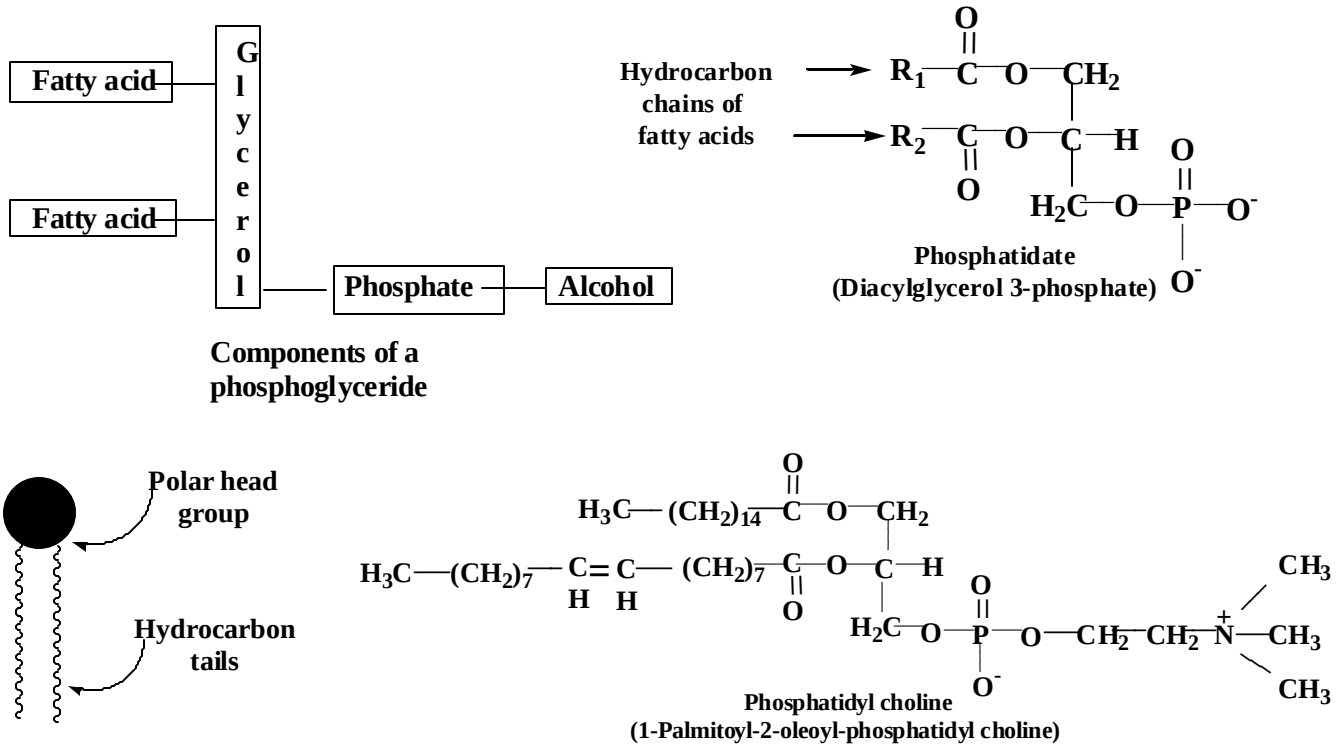


Figure 8: Symbol for phospholipid or glycolipid molecule.

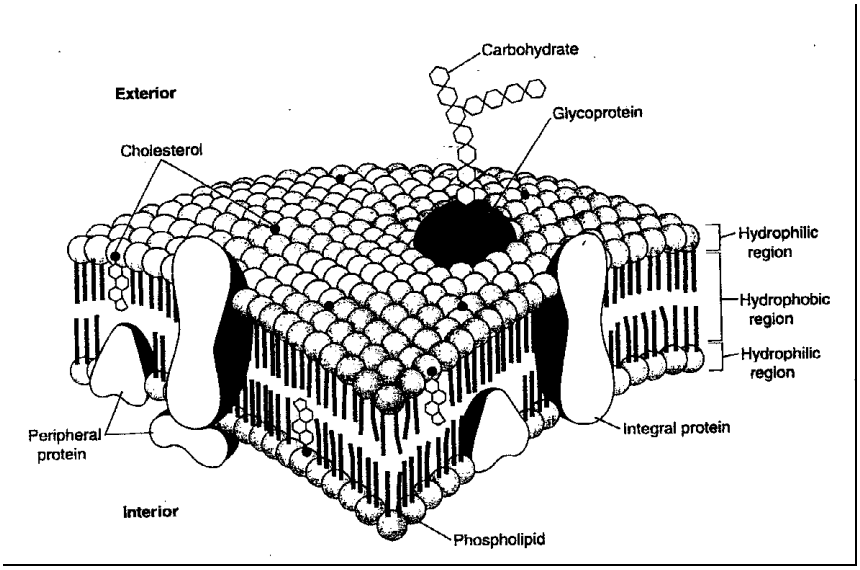


Figure 9: Basic structure of an animal cell membrane.

Triglycerides, also known as fat, are composed of three fatty acids attached to a glycerol backbone:

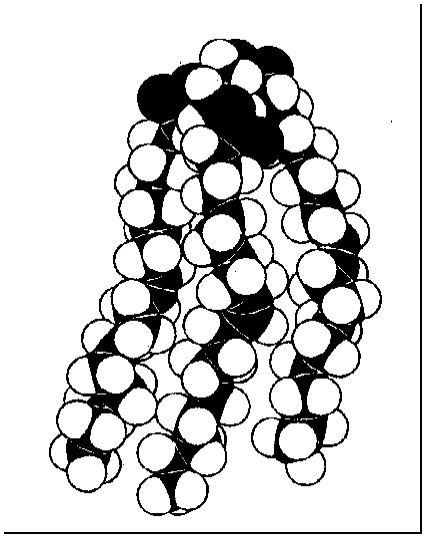


Figure 10: Space-filling model of the trioleate ester of glycerol.

D. Nucleic Acids, RNA and DNA. These are macromolecules made from ribonucleotide and deoxyribonucleotide building blocks, respectively. They form polynucleotide chains that fold into higher order structures; best known is the double helix of DNA.

Table 1: Names of Bases and Their Corresponding Nucleosides[a]

Base	Nucleoside
Adenine (A)	Adenosine
Guanine (G)	Guanosine
Cytosine (C)	Cytidine
Thymine (T)	Thymidine
Uracil (U)	Uridine
Hypoxanithine	Inosine[b]

[a]If the sugar is deoxyribose rather than ribose, the nucleoside has "deoxy" as a prefix (e.g., deoxyadenosine). Nucleotides are given the name of the nucleoside plus mono-, di-, or tri-phosphate (e.g., adenosine triphosphate or deoxyadenosine triphosphate).

[b]The base hypoxanthine is not found in DNA or RNA but is produced during degradation of the purine bases. Its nucleoside, inosine, is produced during synthesis of the purine nucleotides.

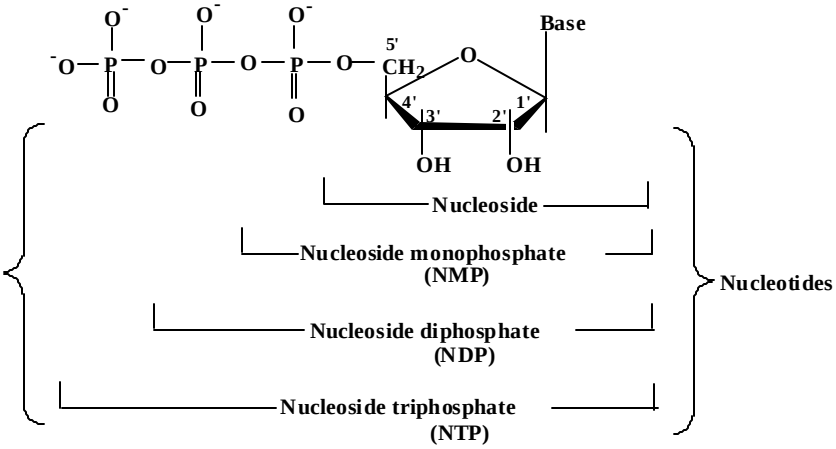


Figure 12: Nucleoside and nucleotide structures. Shown with ribose as the sugar. The corresponding deoxyribonucleotides are abbreviated dNMP, dNDP, and dNTP. N=any base (A, G, C, U, or T).

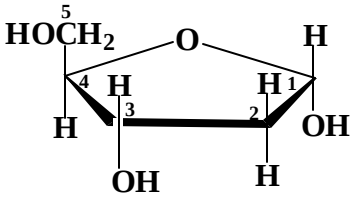


Figure 13: Deoxyribose, the sugar of DNA. The carbon atoms are numbered from 1 to 5. When the sugar is attached to a base, the carbon atoms are numbered from 1' to 5'.

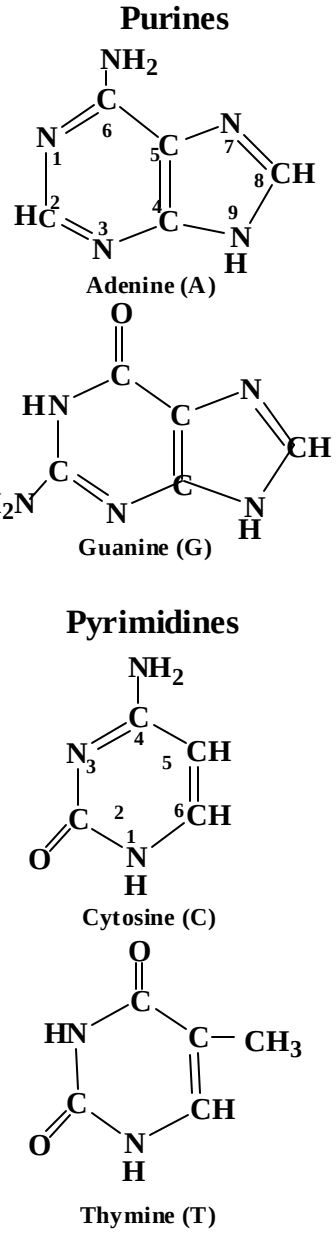


Figure 11: Nitrogenous bases of DNA. Adenine and guanine are purines and cystosine and thymine are pyrimidines.

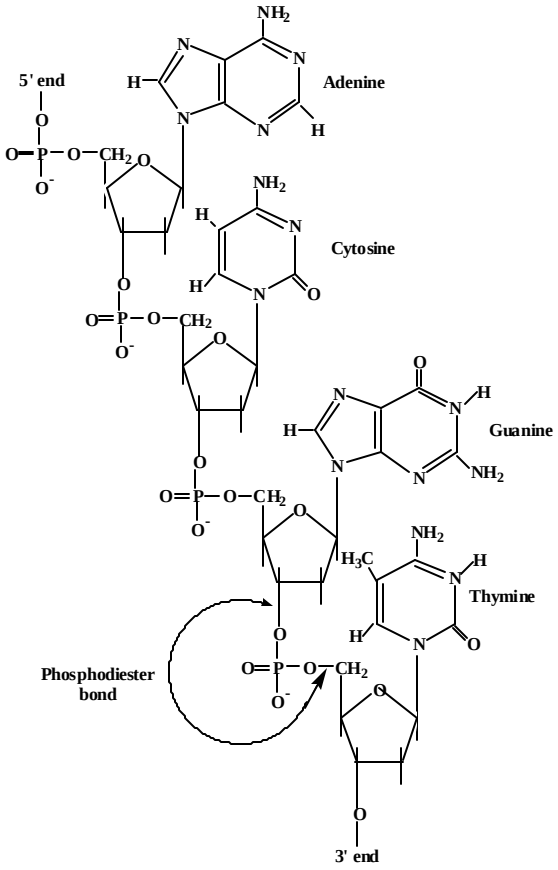


Figure 14: A polynucleotide chain of DNA.

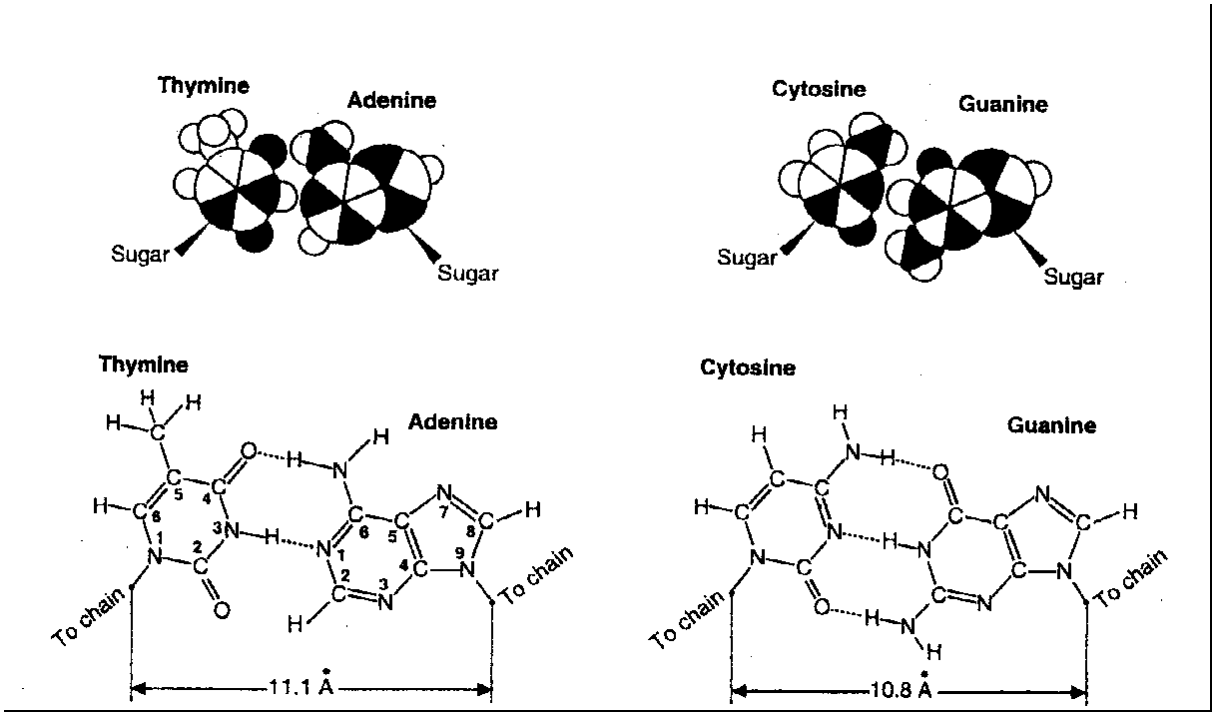


Figure 15: Base pairs of DNA. Note that the pyrimidine bases are "flipped over" from the positions in which they are usually shown. The bases must be in this orientation to form base pairs.

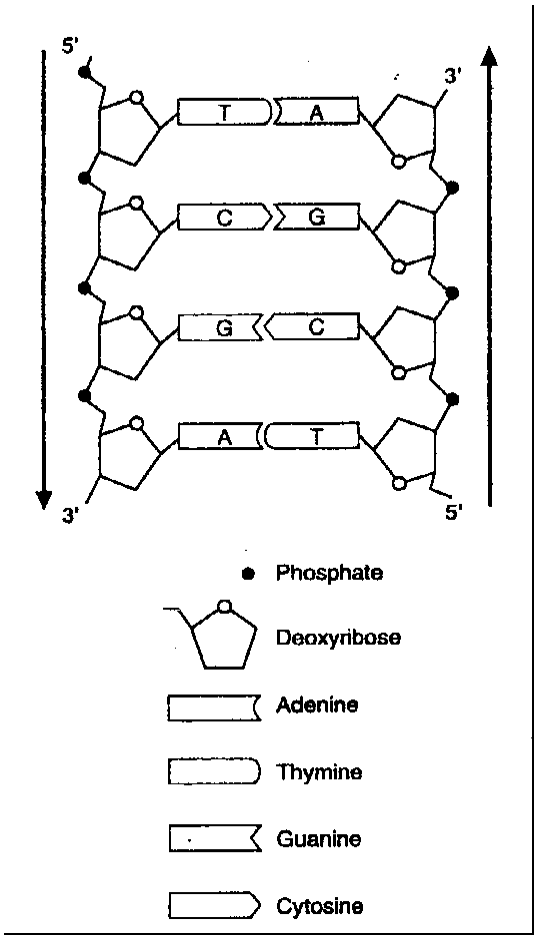


Figure 16: Antiparallel strands of DNA.

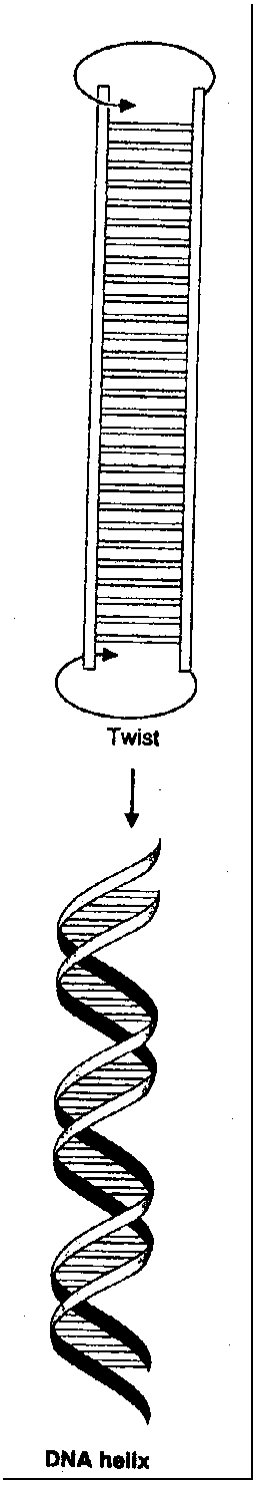


Figure 17: Two DNA strands twisted to form a double helix.

Amino Acids:

Structures & Properties

Learning Objectives for Lectures on:

Amino Acid Structures
Ionization of Amino Acids and Peptides

1. Draw the generic structure of an amino acid and indicate the essential parts: "- carbon, carboxyl group, amino group, "-hydrogen, and side chain.

2. Explain the formalism for determining D- vs L- isomers.

3. Identify each amino acid from its structure, using the full name and 3-letter abbreviation.

4. Describe the characteristic functional group of each side chain.

5. Classify the side chains according to their hydropathy, polarity, and hydrogen bonding potential.

6. Draw the various ionic forms of each amino acid as a function of pH.

7. Estimate the pI's of the individual amino acids or of a dipeptide derived from any pair of them.

8. Explain how hydropathy is determined from oil-water partitioning experiments.

AMINO ACIDS

(Recommended Reading, Chapter 5, Lehninger)

I. Introduction

A. Amino acids are the building blocks of proteins. As described earlier, each amino acid contains a carboxyl group and an amino group attached to a central carbon atom. This motif enables them to form chains by linking together in head to tail fashion through amide bonds. As building blocks they are often referred to as **residues** or **monomeric units**.

B. To create diversity in protein structure and function, amino acids contain a variety of side chains, differing in size, shape, chemical- and physical properties. Twenty kinds of amino acids are commonly found in proteins, and a few more in intermediary metabolism.

There is no perfect way to classify the side chains, but most everyone agrees that the great divide is between polar and nonpolar, as it defines the **hydropathy** (feelings towards water) of the groups. The hydropathy of a residue tells you where you are likely to find it--in the aqueous phase or buried in a hydrophobic environment such as the interior of a protein or the core of a membrane.

Among the polar residues, one may classify them as charged versus uncharged, acidic versus basic. Nonpolar residues are sometimes distinguished as aliphatic or aromatic. As with any classification system, there are some members that don't fit any category very well, or have properties of more than one category.

C. The importance of pH and pK_a. The pH of a solution (a measure of acidity/basicity) and the pK_a (a measure of the strength of an acid or base) determines the state of ionization of the acidic and basic residues. The ionization state in turn has a profound effect on protein structure and function. We will discuss this important concept in a lecture devoted to pH and pK_a.

D. The importance of noncovalent bonds. A variety of noncovalent forces are responsible for the interactions of side chains with each other and with other molecules in solution. These are described as: hydrophobic effects, hydrogen bonds, van der Waals- and electrostatic interactions. These will be discussed in detail in a later lecture.

II. Overview of Structures

- General structure
- Zwitterionic form of "-amino and carboxyl group
- Stereochemistry
- Nomenclature

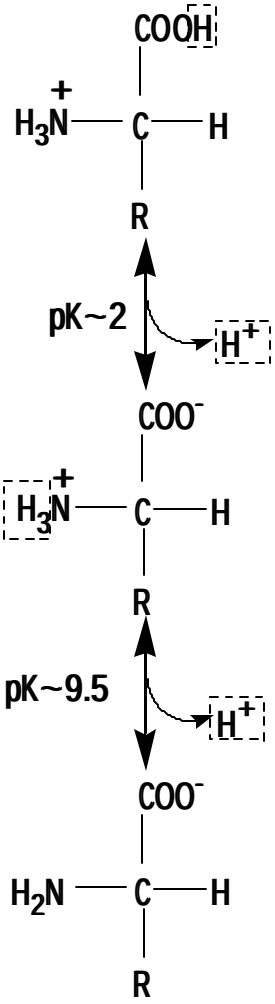


Figure 1: Dissociation of the "-carboxyl and "-amino groups of amino acids. At physiological pH (~7), a form in which both the "-carboxyl and "-amino groups are charged predominates. Some amino acids have ionizable groups on their side chains.

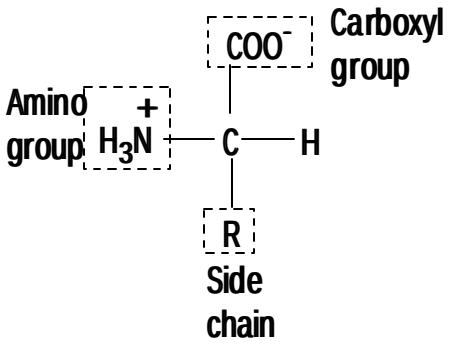
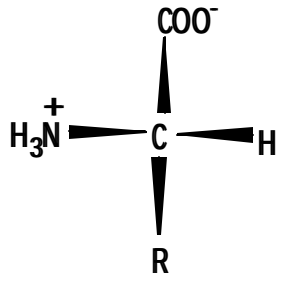
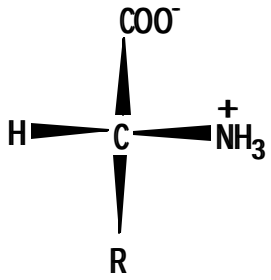


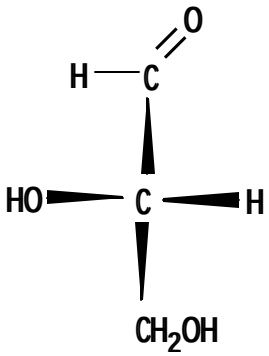
Figure 2: Structure of an amino acid. The "-carbon contains four substituents: an amino group, a carboxyl group, a hydrogen atom, and a side chain (R). Both the amino and carboxyl groups carry a charge at physiological pH.



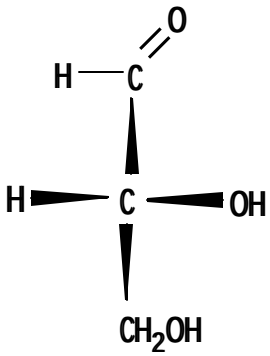
L-Amino Acid



D-Amino Acid



L-Glyceraldehyde



D-Glyceraldehyde

Figure 3: L- and D- amino acids. The L forms are the only ones found naturally in humans. The "-amino group and H-atom come toward the reader, while the "-carboxyl and side chain go away. The L and D forms are mirror images. They cannot be superimposed by rotating the molecule. The reference for the L and D forms are the stereoisomers of glyceraldehyde.

Table 1: Abbreviations for the Amino Acids

	Abbreviations[a]	
Name	**Three Letter**	**One Letter**
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tryosine	Tyr	Y
Valine	Val	V

[a] Three letter abbreviations are generally used. One letter abbreviations are used mainly to list the amino acid sequences of long protein chains.

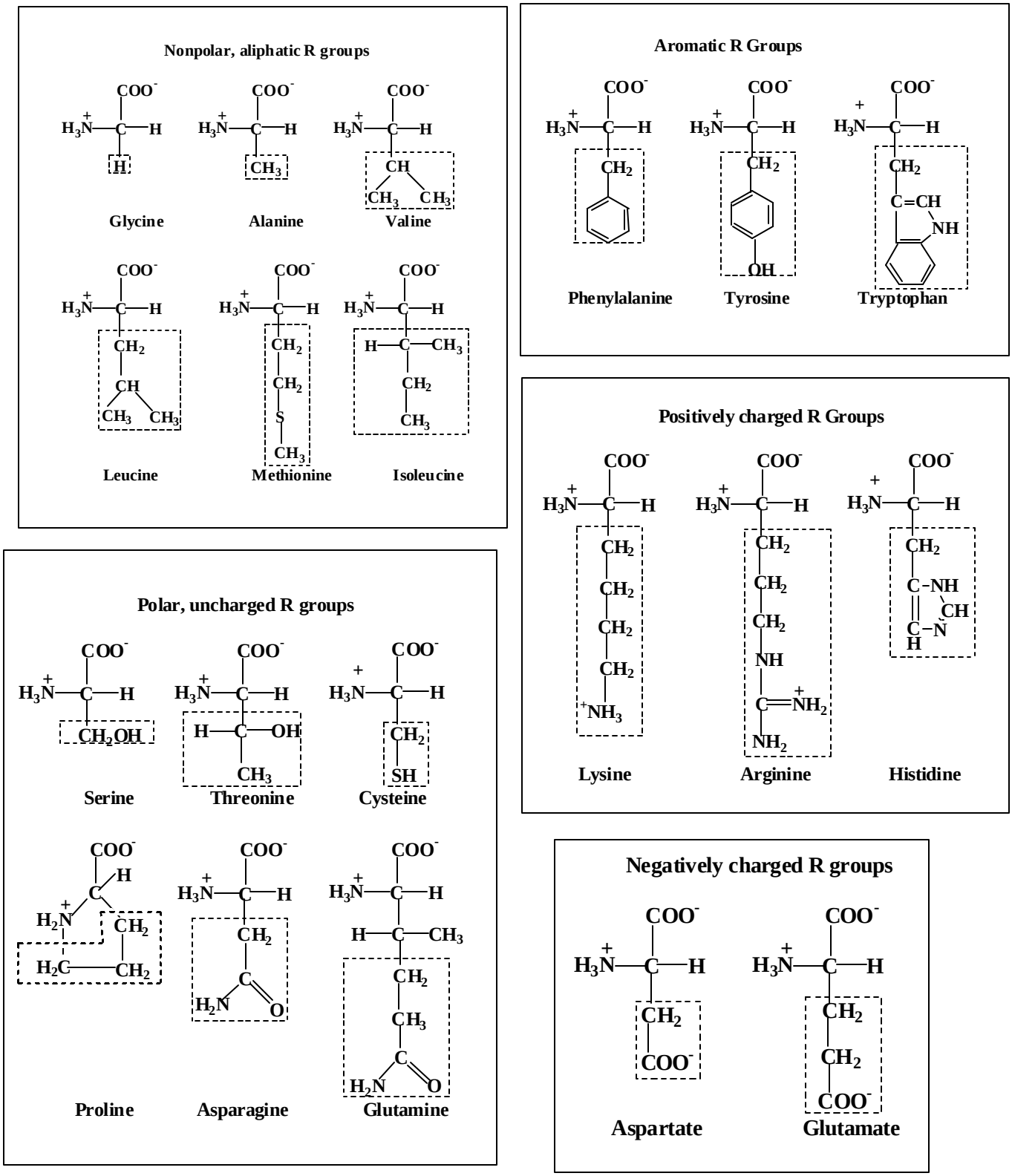
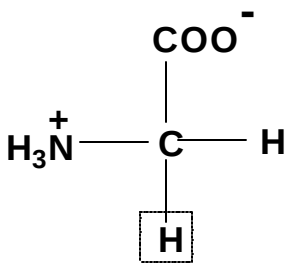


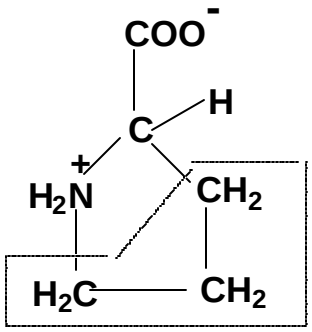
Figure 4: The 20 standard amino acids of proteins. The structural formulas show the state of ionization that would predominate at pH 7.0. The unshaded portions are those common to all the amino acids; the portions shaded are the R groups. Although the R group of histidine is shown uncharged, its pK_a is such that a small but significant fraction of these groups are positively charged at pH 7.0.

III. Specifics



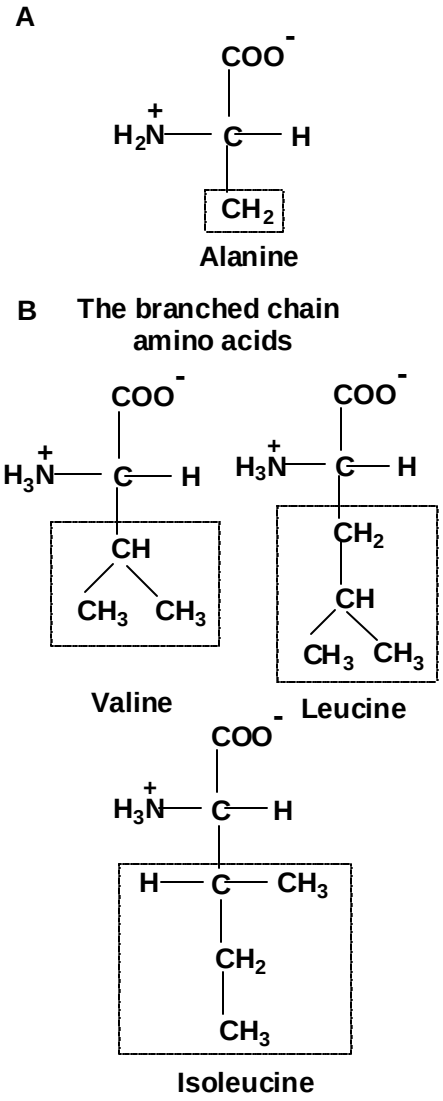
Glycine

Figure 5: Glycine, the simplest amino acid. The "-carbon is not asymmetric because the R group is a hydrogen atom. Therefore, glycine is neither D nor L.



Proline

Figure 6: Proline, a cyclic amino acid. The "-amino nitrogen is incorporated into a ring, which has a rigid structure.



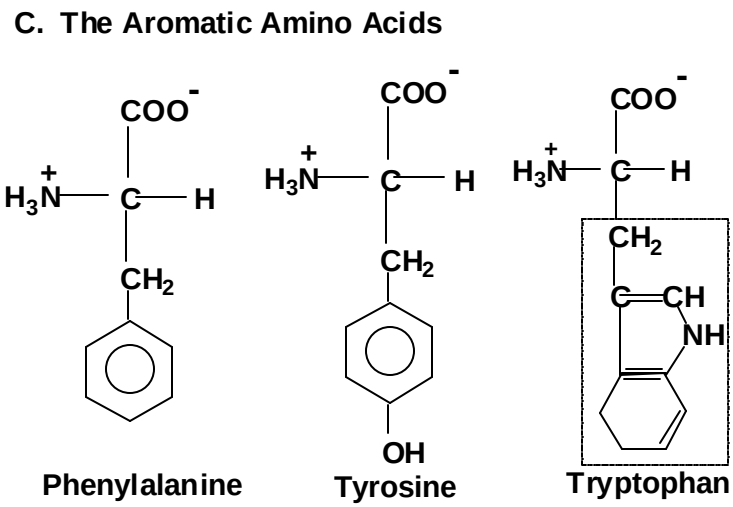
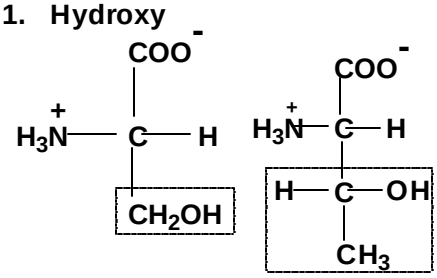


Figure 8: The aromatic amino acids.

Figure 7: Amino acids with aliphatic side chains. A. Alanine. B. Branched chain amino acids (valine, leucine, and isoleucine). The side chains of these amino acids are hydrophobic.

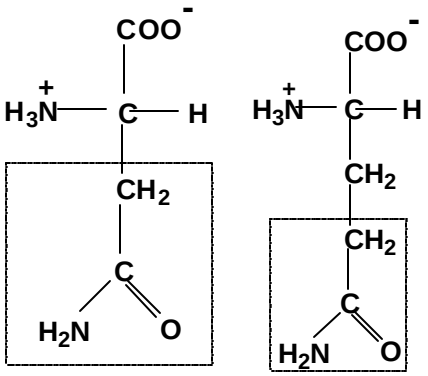
D. Polar , Uncharged Amino Acids

1. Hydroxy



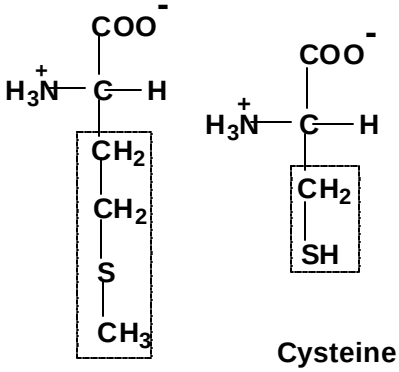
Serine Threonine

2. Amide



Asparagine Glutamine

3. Sulfur-containing



Methionine Cysteine

Figure 9: Hydroxy-, amide-, and sulfur-containing amino acids. Asparagine and glutamine are formed from asparate and glutamate. Serine, threonine, asparagine, and glutamine can form hydrogen bonds. Cysteine can form covalent disulfide bonds.

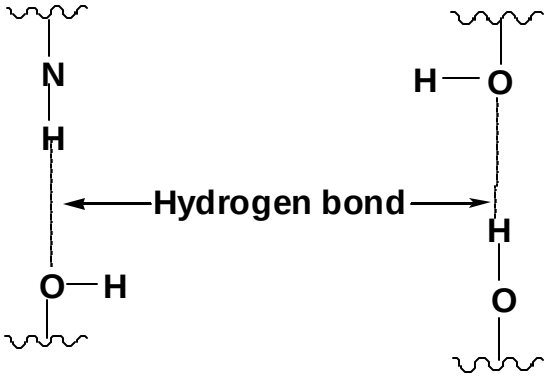
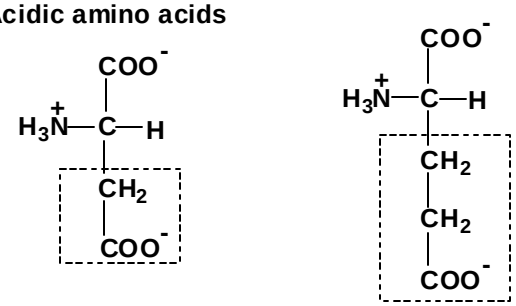


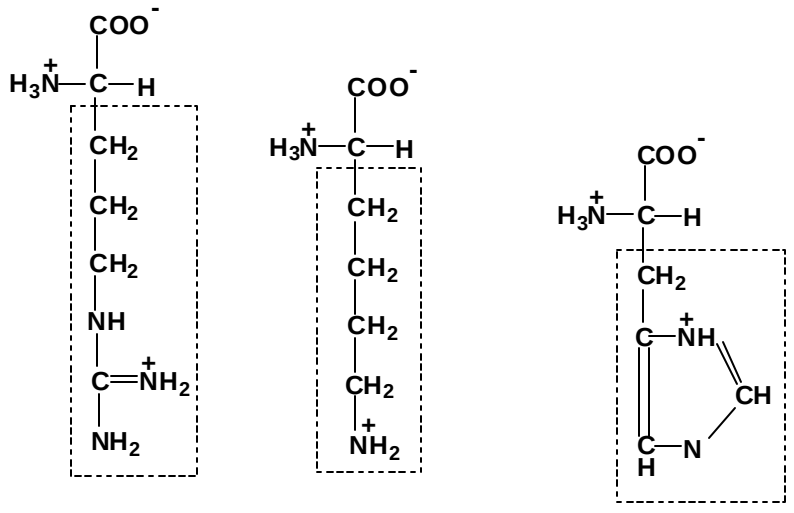
Figure 10: Hydrogen bonds involving nitrogen or oxygen.

E. Acidic amino acids



Asparate Glutamate

F. Basic amino acids



Arginine Lysine Histidine

Figure 11: The acidic and basic amino acids. These amino acids have side chains that carry a charge at physiological pH. The side chains of the acidic amino acids have a carboxyl group that is negatively charged, while the side chains of the basic amino acids have a nitrogen-containing group that is positively charged.

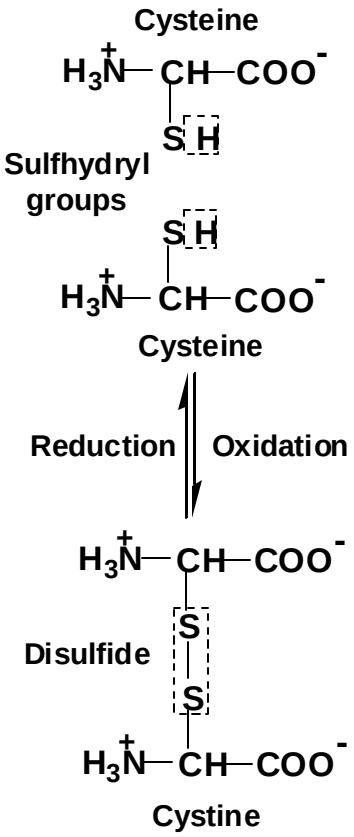


Figure 12: A disulfide bond. Covalent disulfide bonds may be formed between 2 molecules of cysteine or between 2 cysteine residues in a protein. The sulfide compound is called cystine. The hydrogens (with their electrons) of the sulfhydryl groups of cysteine are lost during oxidation.

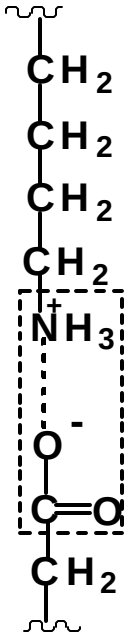


Figure 13: Electrostatic interaction between the positively charged side chain of lysine and the negatively charged side chain of asparate.

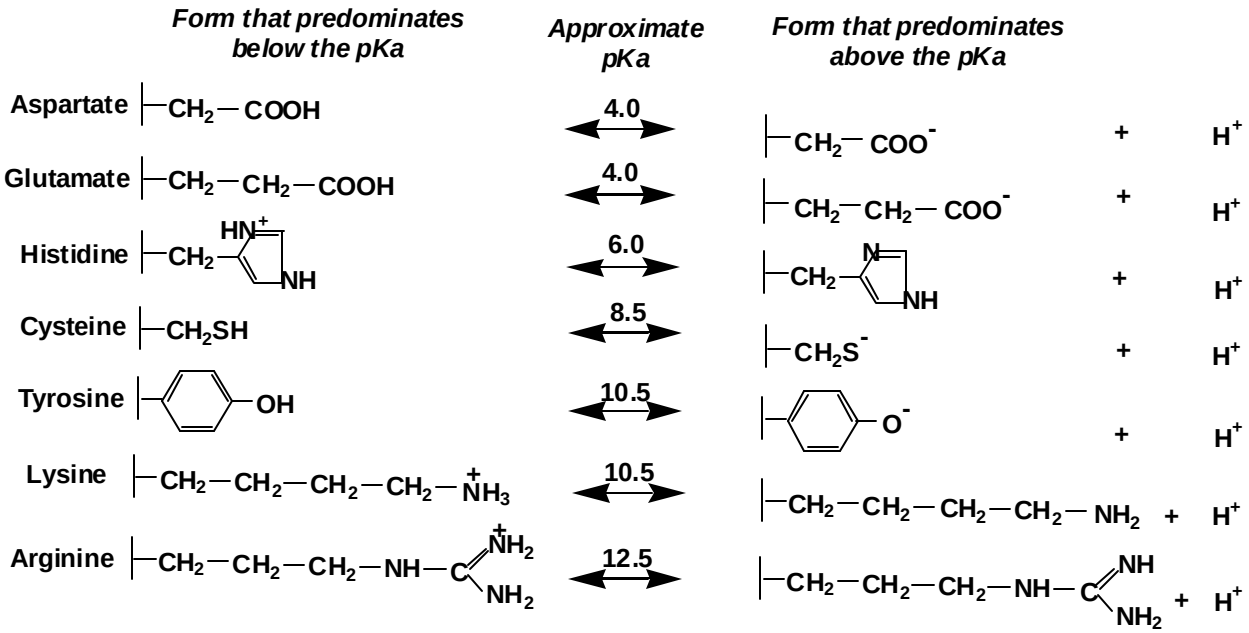


Figure 14: Dissociation of the side chains of the amino acids. As the pH increases, the charge on the side chain goes from 0 to +1, 0 to -1, or +1 to 0. The pK_a is the pH at which half the molecules of an amino acid in solution have side chains that are charged. The net charge on the side chain is therefore +0.5 or -0.5.

IV. Hydropathy

As stated earlier, hydropathy is probably the most important property of the amino acid side chains. There are many ways that hydropathy is determined, experimentally and theoretically. The most straightforward measure of it comes from "water-oil" partition experiments.

An amino acid, typically radiolabelled, is placed in a tube or separatory funnel that contains water and an immiscible organic phase, such as octanol, corn oil, or hexane. The mixture is vigorously shaken, and then the phases are allowed to settle. After complete separation has occurred, the concentration of amino acid in each phase is measured.

The ratio of amino acid concentration in the two phases is called the water-oil partition coefficient:

$$K_p = \frac{[S]_w}{[S]_o}$$

where $[S]_w$ is the concentration of solute in the aqueous phase, and $[S]_o$ is the concentration of solute in the organic (oil) phase. A hydrophobic compound is characterized by $K_p < 1$, and a hydrophilic compound by $K_p > 1$.

Tables of partitioning data are usually expressed as log K_p, or as the free energy of transferring the solute from oil to water, ${}^aG_{o->w}$, where, under standard conditions

$$^aG_{o6w} = -1.36 \log K_p$$

The tables usually show the hydropathies of the amino acid side chains rather than the hydropathies of the free amino acids. Side chain hydropathies have been calculated by subtracting the transfer free energy of the $^+H_3N\text{-}CH\text{-}COO^-$ moiety from the transfer free energies of each amino acid.

A hydropathy table of side chain transfer free energies is shown on the following page (Table 2). Hydrophobic side chains have positive values, whereas hydrophilic side chains have negative values. The corresponding K_p values range from 2.41×10^{-4} for leucine to 9.34×10^{10} for arginine.

The K_p's for ionizable side chains such as those of asp, glu, lys, and his reflect the partitioning of all ionic forms between water and the organic phase at pH 7.0. Because charged species such as COO^- are far more hydrophilic than uncharged polar species such as COOH, the K_p's vary dramatically with pH owing to the change in ratio of charged to uncharged forms; asp and glu are far less hydrophilic at low pH, whereas his and lys are far less hydrophilic at high pH.

It is also important to be aware that arg, lys, tyr, and trp are amphipathic, with hydrophilic and hydrophobic domains. So, for example, the aromatic ring of tyr may be buried in the interior of a membrane, with the -OH group in the aqueous phase. The hydropathy index does not reflect this property.

Not shown is the hydropathy of the polypeptide backbone itself, which is about the same as that of amides (hydrophilic).

Table 2: Free energies of transfer of amino acid side chains to water from cyclohexane

Cyclohexane ⇌ H_2O

Side chain of	kcal/mol
Leu	4.92
Ile	4.92
Val	4.04
Pro	3.58
Phe	2.98
Met	2.35
Trp	2.33
Ala	1.81
Cys	1.28
Gly	0.94
Tyr	-0.14
Thr	-2.57
Ser	-3.40
His	-4.66
Gln	-5.54
Lys	-5.55
Asn	-6.64
Glu	-6.81
Asp	-8.72
Arg	-14.92

Free energies of transfer at 25°C and pH 7.0

For our purposes, a semi-quantitative hydropathy scale will suffice, as shown below:

Table 3: Semi-quantative table of side chain hydropathies. The plus, minus, and zero symbols indicate hydrophobic, hydrophilic, and "hydro-apathetic" respectively.

Amino Acid	Hydropathy
gly	0
ala, val, leu, ileu, pro, phe	+
ser, thr	-
tyr	0
asp, glu	-
asn, gln	-
cys	0
met	+
lys	-
arg	-
his	-
trp	+
Peptide bond	-

V. Physical basis of hydropathy

We will be discussing the physical basis of hydropathy in more detail in a subsequent lecture. For the time being we can make use of a venerable rule from chemistry: "like dissolves like." Water is a highly polar molecule with an extremely high capacity to form hydrogen bonds. Therefore, it is far better than an organic solvent for dissolving hydrogen-bonding solutes. The high polarity also makes it an excellent solvent for ions.

Polarity refers to a skewed distribution of electrons in a bond, so that one atom bears a partial negative charge whereas the other bears a partial positive charge. Most of the polar bonds we will encounter contain oxygen or nitrogen:

δ^--O-H δ^+

δ^-N-H δ^+

δ^+-C-O-δ^-

δ^+-C=Oδ^-

Summary

Table 3: Semi-Quantitative Properties of Amino Acids

Amino Acid	pK_{a1} (Carboxyl)	pK_{a2} (Amino)	pK_{a3} (R-Group)	Hydropathy *
Nonpolar aliphatic	2.0	9.5	–	
Glycine				0
Proline				+
Alanine				+
Leucine				+
Valine				+
Isoleucine				+
Aromatic	2.0	9.5		
Phenylalanine				+
Tyrosine			10	0
Tryptophan				+
Polar, uncharged	2.0	9.5	–	
Cysteine			8.0	0
Methionine				+
Threonine				-
Serine				-
Asparagine				-
Glutamine				-
Charged-negative	2.0	9.5	4.0	
Aspartate				-
Glutamate				-
Charged-positive	2.0	9.5		
Histidine			6.0	-
Lysine			10.5	-
Arginine			12.2	-

* +, -, 0 mean hydrophobic, hydrophilic, “hydroapathetic” respectively.

pH & pK_a:

Acids, Bases, & Buffers

pH & pK_a:
ACIDS, BASES, AND BUFFERS

(Recommended reading: Lehninger pp. 95-111)

I. Introduction

Every protein is designed to operate within a narrow pH range, where its structure is most stable and its biological activity is optimally "tuned." For this reason primarily, the pH of cells, organelles, and the aqueous compartments of the body are tightly controlled.

Example: The normal pH of blood is 7.4 and is maintained within the narrow range of 7.37-7.43. The clinical conditions of **alkalosis** and **acidosis** are manifest at higher and lower pH, respectively. Death occurs when the deviation from normal reaches pH 7.8 on the alkaline side and 6.8 on the acid side.

The clinical symptoms resulting from these pH abnormalities can always be tracked back to the malfunction of one or more proteins such as the oxygen carrier hemoglobin.

In this lecture we will review acid-base chemistry, introduce the concept of buffering, and show how these principles apply to the maintenance of blood pH.

II. A brief review of aqueous solutions and concentration units

Most of the reactions in the body take place in aqueous solution. (The rest take place in membranes or at the interface between phospholipid monolayers and the aqueous phase.) Generally, the most important measure of a dissolved substance is its concentration.

Molarity is the concentration scale most useful to biochemists. A one molar solution, written as 1 M, contains one mole of solute dissolved in one liter of solution. To give you an intuitive feel for this concentration, consider the protocol for making a 1 M solution of sucrose (ordinary table sugar):

> The molecular weight of sucrose is 342 g/mole. Therefore a 1 M solution is 342 g/l, which is equivalent in "kitchen units" to about 1.66 cups of sugar per quart. This is a very concentrated solution! One molar solutions of strong acid and alkali bases will burn the skin.
>
> A typical 12 oz soft drink is 0.31 M sucrose (38 g/355 ml).
>
> Rarely are solutes found at such high molarities in aqueous biological compartments. Molecules such as simple sugars, amino acids, water soluble vitamins, and so on, have concentrations in the range of 0.001 to 0.01 M. For example, the normal level of blood glucose is 0.005 M, or 5 millimolar (5 mM). The concentrations of macromolecules such as proteins are typically in the **micro**molar, **nan**omolar, or even **pico**molar range.

Table: Frequently Used Prefixes in the Metric System

Prefix	Symbol	Value Relative to the Unprefixed
femto	f	10^{-15}
pico	p	10^{-12}
nano	n	10^{-9}
micro	μ	10^{-6}
milli	m	10^{-3}
centi	c	10^{-2}
deci	d	10^{-1}
unit		10^{0}
kilo	k	10^{3}

For the purpose of identifying what is clinically normal or abnormal, physicians prefer to use units of milligrams per deciliter (mg/dL), where a deciliter is 100 ml. A typical fasting blood glucose level is 90 mg/dL, or 90 **milligram percent**.

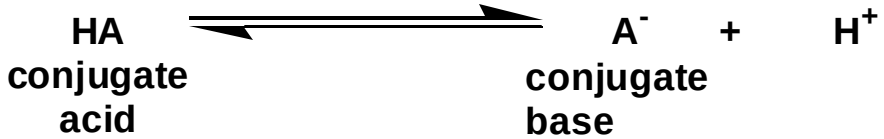
The concentration of salts in the blood is about 0.15 M. The main contributor is NaCl, at 0.12 M. To make a 0.12 M solution of NaCl, add 7 g to 1 liter of water.

(In case you don't have a feel for grams–an aspirin tablet is about 0.375 g. A teaspoon of sugar is 4 g.)

III. Definitions and Descriptions

A. Bronsted Concept of Conjugate Acid-Conjugate Base Pairs

We define an acid as a substance that donates protons (hydrogen ions, H^+) and a base as substance that accepts protons. When the acid loses a proton, a base is produced. The original acid and resulting base are referred to as a **conjugate acid-conjugate base** pair.



B. Ionization of a strong acid

A "strong" acid is a substance such as HCl that ionizes almost 100% in aqueous solution to produce H_3O^+ and Cl^-

$$HCl + H_2O \longrightarrow H_3O^+ + Cl^-$$

H_3O^+, the hydronium ion, is the actual form of the hydrogen ion in solution. For simplicity we write the reaction as:

$$HCl \longrightarrow H^+ + Cl^-$$

A "strong" base is a substance that ionizes almost completely in water to yield OH^- ions. The most familiar examples are sodium and potassium hydroxides.

$$KOH \longrightarrow K^+ + OH^-$$

IV. Ionization of water

Water ionizes very slightly to produce both a proton and a hydroxide ion.

$$HOH \rightleftharpoons H^+ + OH^-$$

The equilibrium lies far to the left. In pure water, $[H^+] = [OH^-] = 10^{-7}$ M. Since the molarity of HOH is about 55.6 M (see if you can show this), the ratio of dissociated water to undissociated water is only $(10^{-7}/55.6) = 1.8 \times 10^{-8}$; less than 1 water molecule per 60,000,000 is dissociated.

The equilibrium expression for the ionization is

$$K_d = \frac{[H^+][OH^-]}{[HOH]} = \frac{(10^{-7})(10^{-7})}{(55.6)}$$

$$= 1.8 \times 10^{-16} M$$

where K_d is the **ionization constant** or **dissociation constant.**

Addition of acids or bases to water changes the concentrations of hydrogen ions and hydroxide ions reciprocally (if one goes up by a factor of ten, the other goes down by a factor of ten) but the relative change in [HOH] is negligible so we treat [HOH] as a constant; we define a new dissociation constant, K_w, by multiplying each side of the equilibrium expression by the factor 55.6:

$$K_d(55.6) = [10^{-7}][10^{-7}]$$

$$K_w = 10^{-14} = [H^+][OH^-]$$

K_w is often referred to as the **ion product** of water. It is a very handy constant because it quantitatively relates the reciprocal concentrations of hydrogen ion and hydroxide ion.

Example: If the hydrogen ion concentration is 10^{-3} M, what is the hydroxide ion concentration?

Answer: 10^{-11} M.

V. Acidity, Basicity, and the pH scale

The acidity or basicity of a solution can be expressed as either the hydrogen ion concentration or the hydroxide ion concentration, but usually $[H^+]$ is preferred. When $[H^+] = [OH^-]$, the solution is neither acidic nor basic, but **neutral**. Since the ion product of water is 10^{-14} neutrality occurs when the concentration of both species = 10^{-7} M. When $[H^+]$ is $> 10^{-7}$ M, the solution is acidic; when $[H^+] < 10^{-7}$, the solution is basic (alkaline).

To avoid the inconvenience of using long numbers or scientific notation to express $[H^+]$, a shorthand notation has been devised, called **pH**, defined as the negative logarithm of the hydrogen ion concentration:

$$pH = -\log[H^+] = \log 1/[H^+]$$

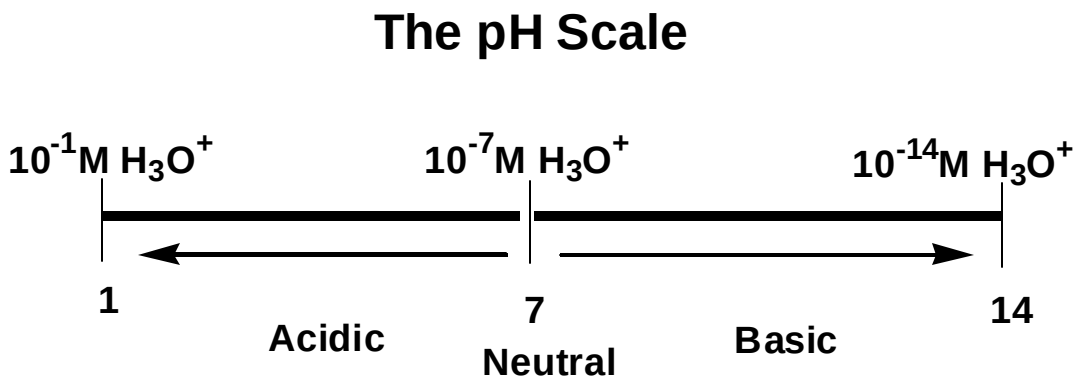
(A rule of logs: the negative log of a number equals the positive log of the reciprocal of the number.)

Sample problem: What is the pH of pure water at neutrality?

Solution: Since $[H^+] = 10^{-7}$, pH = $-\log 10^{-7} = -(-7) = 7$

The highest concentrations of acids or bases that we usually consider are 1.0 M, corresponding to pH values of 0 and 14 respectively (work this out for yourself). Therefore, the usable **pH scale** is 0 to 14, with a midpoint at pH = 7. Some typical values are shown below:

Solution	**pH**	**H_3O^+ Concentration (M)**
Gastric juice	1-2	10^{-1} to 10^{-2} M
Cola drink	3	10^{-3} M
Urine	5-8	10^{-5} to 10^{-8} M
Saliva	6.4	4 X 10^{-7} M
Pure water	7.0	10^{-7} M
Blood	7.4	4 X 10^{-8} M



VI. Effect of adding strong acids or bases on the pH of water

The pH of an aqueous solution is conveniently measured with--guess what--a pH meter. Therefore, the effect of adding an acid or base to a solution is easily determined experimentally. Nevertheless, it is important to be able to calculate the approximate effect of adding acids and bases to solution in order to understand the concept of buffering (later section).

Sample problem: If you add 1ml of 1 M HCl to 999 ml of water, what will be the final pH?

Solution: Calculate a) the final volume and, b) the moles of HCl contained in that volume.

Calculating the final volume is easy: 999 ml water + 1 ml acid = 1 liter solution.

To calculate the moles of HCl added, let's use the millimolar scale. A handy thing to remember is that a 1.0 M solution contains 1mmole/ml. Consequently, when you add 1ml of 1M HCl, you are adding 1mmole of HCl.

So–the final solution contains 1mmole HCl in a volume of 1 liter; $[H^+]$ = 1mmole/l liter = 10^{-3} M. The pH is -log 10^{-3} = -(-3) = 3.

The method we used to solve this problem contains some hidden simplifying assumptions. One of them is, that all the H^+ comes from the added HCl, which neglects the fact that water itself contributes 10^{-7} M protons via dissociation: HOH <==> H^+ + OH^-. The small contribution by water to [H^+] can be safely ignored if the [HCl] is >>10^{-7} M, but not when [HCl] is comparable to or less than 10^{-7} M. Keep this in mind when we discuss the neutralization of acids with bases.

The effect on pH of adding a strong base to pure water can be calculated in similar fashion to calculating the effect of adding strong acid. (See problems at end.)

VII. Effect of titrating a strong acid with a strong base

When a strong base such as KOH is added to a solution of strong acid, each molecule of OH^- reacts (essentially) irreversibly with a molecule of H^+ to produce a molecule of HOH.

Example: if 2 mmoles of KOH is added to a solution that contains 5 mmoles of HCl, the new solution will contain a net excess of 3 mmoles H^+. (The water produced via neutralization is negligible compared to the amount that is already present, so we ignore the small increase when calculating the final H^+ concentration.)

Sample problem:

If you add 50 ml of 0.02 M KOH to 500ml of an HCl solution, pH 2.5, what will be the final pH?

Solution:

(1) Calculate the moles of reactants

KOH: (0.02 mmoles/ml)(50 ml) = 1.0 mmoles KOH

HCl: This requires two steps

(a) [H^+] = antilog (-2.5) = 3.16 X 10^{-3} moles/l

(b) (3.16 X 10^{-3} mmoles/ml)(500 ml)=1.58 mmoles HCl

(2) Calculate the moles of products

1.58 mmoles HCl +1.0 mmoles KOH –> (1.58-1) = 0.58 mmoles HCl in excess.

(3) Calculate the final volume

(50 ml KOH + 500 ml HCl)=550 ml final volume

(4) Calculate the final concentration

0.58 mmoles HCl/550 ml)=1.05 X10^{-3} M HCl

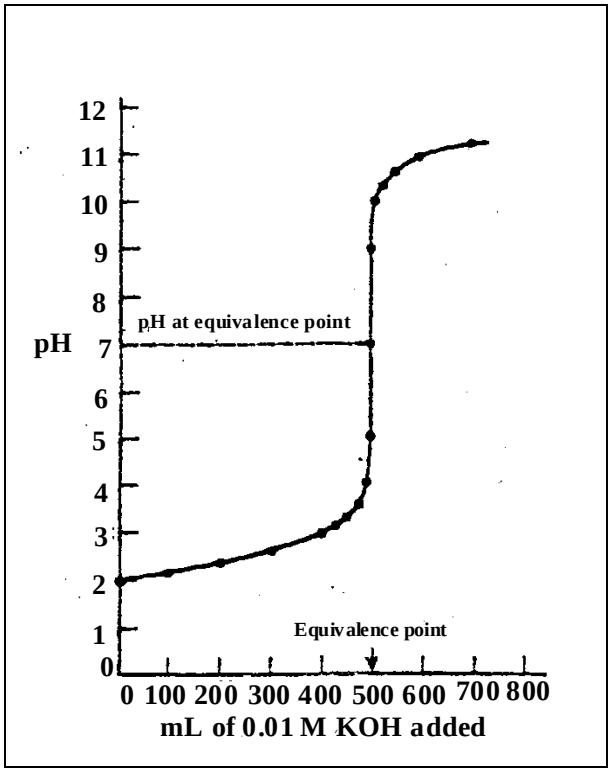
(5) Calculate the final pH

pH = -log(1.05 X 10^{-3}) = 2.98

If in the Example, 5 mmoles KOH had been added to the 5 mmoles HCl, the excess HCl would have been exactly neutralized. Our simplifying assumption described earlier would lead us to conclude that the final $[H^+]$ should be zero. This conclusion is wrong because H_2O itself contributes H^+ via dissociation. Therefore, when the HCl is exactly neutralized, the $[H^+]$ is 10^{-7}, not zero, and the pH is 7.

The graph below is a **titration curve**, showing what happens when KOH (0.01M) is added in increments to a solution of 0.01M HCl.

Figure 1: Titration of a strong acid (e.g., HCl) with a strong base (e.g., KOH).



Notice that the starting pH is 2 because of the $[H^+]$. Over most of the titration, the pH does not change drastically because pH is a log scale of $[H^+]$. For example a 10-fold reduction in [HCl], corresponding to 90% neutralization raises the pH just one unit, from 2 to 3; a 100-fold reduction in [HCl], corresponding to 99% neutralization, raises the pH two units, from 2 to 4. But the pH shoots up as you near the equivalence point. Why? Because the amount of excess H^+ that remains is small compared to the amount of KOH contained in each addition of base.

This graph illustrates that in the physiological range of pH, near 7, addition of a very small amount of acid or base does change the pH drastically.

> For example, adding just one micromole (10^{-6} mole) of HCl to 1 liter of pure water increases the $[H^+]$ approximately 10-fold--from 10^{-7} to 10^{-6}--which reduces the pH approximately one whole unit, from 7 to 6; similarly, adding just one micromole of KOH increases the $[OH^-]$ ten fold, causing the pH to rise approximately one whole unit, from 7 to 8.

We say that pure water has very little **buffering capacity**. This term will be defined rigorously later, but for now think of it semi-quantitatively as a measure of the resistance to pH change upon addition of acid or base. We will find that in the presence of a suitable buffer, a thousand times as much acid or base would have to be added to see such a change in pH.

VIII. Ionization of weak acids

A. Monoprotic acids. (These acids have just one hydrogen atom to donate.)

Acetic acid (the "active" ingredient of vinegar) is a typical weak acid. A 0.1 M solution is dissociated to the extent of only about 1%, producing a hydrogen ion concentration of about 10^{-3} M and a corresponding pH of 3. (In contrast, a 0.1 M solution of HCl produces a hydrogen ion concentration of 0.1 M and pH=1.)

The dissociation reaction can be described by the simplified equation:

$$HAc \rightleftharpoons H^+ + Ac^-$$

$$K_a = \frac{[H^+][Ac^-]}{[HAc]} \qquad \text{Eqn VIII.1}$$

where HAc (acetic acid) is the conjugate acid and Ac^- (acetate ion) is the conjugate base. The acid dissociation constant for this reaction, $K_a=1.8 \times 10^{-5}$, indicates that the equilibrium lies far to the left.

Notice that the smaller K_a is, the weaker is the acid and vice versa.

Just as we found it convenient to define a $[H^+]$ scale in terms of the pH, we will find it convenient to define a scale of acid dissociation constants in terms of **pK_a**, where:

$$pK_a = -\log K_a$$

For acetic acid, the $pK_a = -\log (1.8 \times 10^{-5}) = 4.74$.

A smaller value of K_a corresponds to a higher value of pK_a and vice versa. Consequently, acid strength is inversely related to pK_a.

Sample problem: The K_a's of acetic acid and lactic acid are 1.8×10^{-5} and 1.38×10^{-4} respectively. What are the pK_a's? Which one is the strongest acid?

Answer: pK_a (acetic) = $-\log 1.8 \times 10^{-5} = 4.74$; pK_a (lactic) = $-\log 1.38 \times 10^{-4} = 3.86$. Lactic acid is stronger.

Having defined the pH and pK_a scales, we will find that Eqn. VIII.1 is more convenient to use if we transform it as follows:

$$-\log K_a = -\log \frac{[H^+][A^-]}{HAc}$$

$$-\log K_a = -\log [H^+] - \log \frac{[A^-]}{[HAc]}$$

$$pK_a = pH - \log \frac{[A^-]}{[HAc]}$$

$$pH = pK_a + \log \frac{[A^-]}{[HAc]}$$

This relationship, written for the general case below, is known as the **Henderson-Hasselbalch** equation, and is used extensively to solve problems dealing with weak acids, weak bases, and buffers. It should be memorized:

$$pH = pK_a + \log \frac{[\text{conjugate base}]}{[\text{conjugate acid}]} \qquad \text{Eqn. VIII.2}$$

B. Polyprotic acids

Acids such as acetic and hydrochloric are called monoprotic because they can donate only one proton. Other acids, such as sulfuric (H_2SO_4) are polyprotic, meaning they can donate more than one. Two important inorganic acids in this class are phosphoric (H_3PO_4) and carbonic (H_2CO_3). The ionization of phosphoric will be discussed below and the ionization of carbonic under the section on blood gases.

The ionization of phosphoric acid is characterized by three pK_a's, each corresponding to the loss of one proton.

$$H_3PO_4 \rightleftharpoons H_2PO_4^{-1} + H^+ \; ; \; pK_a(1) = 2.1$$

$$H_2PO_4^{-} \rightleftharpoons HPO_4^{-2} + H^+ \; ; \; pK_a(2) = 6.7$$

$$HPO_4^{-2} \rightleftharpoons PO_4^{-3} + H^+ \; ; \; pK_a(3) = 12.3$$

Three conjugate bases are produced: $H_2PO_4^{-1}$, HPO_4^{-2}, and PO_4^{-3}, which are called monobasic–, dibasic–, and tribasic phosphate, respectively. The pK_a's are sub-labeled as 1, 2, 3, in order of strongest acid to weakest.

C. Weak bases

We have defined an acid as a substance that donates a proton and a base as a substance that accepts a proton. Just as K_a is a measure of acid strength, there is an analogous scale of K_b to measure base strength. However, few people use the K_b scale, mainly because it is not necessary. The K_a-pK_a scale can be used indirectly as a measure of basicity. How? By taking advantage of the fact that as a series of bases increases in base-strength, their accompanying conjugate acids decrease in acid-strength. Consequently, the pK_a of a conjugate acid is an indirect measure of the basicity of the parent base.

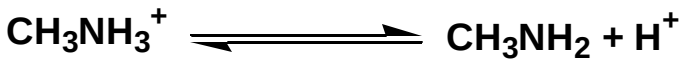
Example: Ammonia is a weak base. When added to water, the pH rises because of the reaction to form ammonia ion:

The equilibrium lies well to the right, with a $K_{eq} = 2.5 \times 10^9$.

If we write the reaction in reverse, it has the same form as an acid dissociation:

The equilibrium constant for the reaction in this direction is formally an acid dissociation constant, equal to the reciprocal of K_{eq}. Hence, $K_a = 1/K_{eq} = 3.98 \times 10^{-10}$ and $pK_a = -\log 3.98 \times 10^{-10} = 9.4$. Compared with acetic acid, K_a for NH_4^+ is more than 5 orders of magnitude smaller and pK_a is more than 5 units higher, both reflecting the fact that NH_4^+ is a much weaker acid than HAc. (And also, that NH_3 is a much stronger base than Ac^-).

Sample problem: The pK_a for the methylammonium ion is 10.6.



Is methylamine a stronger or weaker base than ammonia?

Answer: Since the pK_a of the methylammonium ion is higher than the pK_a for the ammonium ion ($pK_a = 9.4$), methylammonium ion is a weaker acid than ammonium ion; therefore methylamine is a stronger base than ammonia.

NOTE CAREFULLY. Biochemists generally refer to the "pK_a of an amino group," when they really mean the pK_a of the conjugate acid of the amine.

IX. Titration of weak acid–production of a buffer

When a strong base such as KOH is added to a solution of weak acid, such as acetic, each molecule of OH^- neutralizes one molecule of HAc, producing one molecule of HOH and one molecule of Ac^-. As a short cut to calculating the concentrations of conjugate acid and conjugate base at equilibrium, we often pretend that the reaction takes place by the direct action of OH^- on HAc:

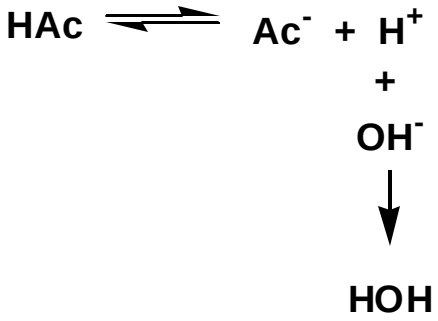
$$HAc + OH^- \rightarrow HOH + Ac^-$$

We also regard the reaction as essentially irreversible. With these assumptions, it is easy to calculate how much HAC remains and how much Ac^- is produced after an addition of KOH.

Example: If you start with 5 mmoles of HAc and add 3 mmoles of KOH to the solution, you will produce a mixture of 3 mmoles Ac^- and 2 mmoles HAc. (Notice that we also ignored the small concentration of Ac^- that was initially present.)

Nevertheless, the physical process is more complex than it appears and illustrates one of the most important concepts in biochemistry: the principle of mass action, a.k.a., the Principle of LeChatelier.

Before the KOH is added, an equilibrium exists among HAc, A^-, and H^+. Addition of OH^- temporarily perturbs the equilibrium by reacting with H^+. Within milliseconds, a new equilibrium is established as the reaction is "pulled" to the right by removal of a product. This is a pair of **coupled reactions**:



After each addition of base, we want to be able to calculate the pH at the new equilibrium position. The approximate solution can be obtained with the Henderson-Hasselbalch equation as shown in the following example:

Sample problem: Start with 1 liter of 1 M acetic acid. Add 20 ml of 1 M KOH. What will be the final concentration of $[H^+]$, [HAC], and $[Ac^-]$?

Solution:

(1) Calculate the moles, or mmoles of reactants

HAc: (0.1mmoles/ml)(1000 ml)=100 mmoles HAc

KOH:(1mmole/ml)(20 ml)=20 mmoles KOH

(2) Calculate moles of products

100 mmole HAc + 20 mmol KOH –> 80 mmol HAc + 20 mmol Ac^-

(3) Calculate the total volume

(1000 ml acetic acid solution + 20 ml KOH solution) = 1020 ml total volume

(4) Calculate the concentration of products

[HAc] = (80 mmoles/1020 ml)=0.0784 M HAc

$[Ac^-]$ = (20 mmoles/1020 ml)=0.0196 M Ac^-

(5) Calculate pH from the H-H equation

$pH = pK_a + \log([Ac^-]/[HAc])$

$pH = 4.74 + \log(0.0196/0.0784)$

$= 4.1$

(6) Calculate $[H^+]$ from pH according to formula as shown

$[H^+]$ = antilog (-pH)=7.94×10^{-5} M

A titration curve for a 0.1M solution of a carboxylic acid with a pK_a=5 is shown below:

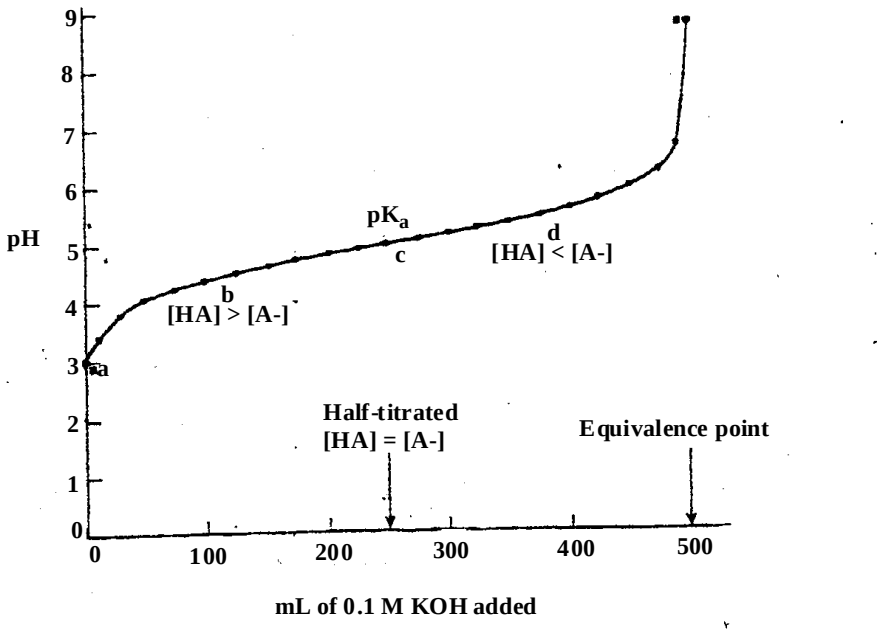


Figure 2: Titration of a weak monoprotic acid, HA, with a strong base (e.g., KOH). pK_a = 5.0; [HA] at the start = 0.1M.

Notice that addition of small amounts of base causes the pH to rise abruptly at the beginning and end of the titration, but only gradually through the middle of it. In the long "plateau" region, the solution is relatively resistant to pH change upon addition of base.

This plateau is reminiscent of the one we saw earlier for the titration of a strong acid with a strong base. The important difference is that with the weak acid, the plateau occurs at a much higher pH, in this case approaching the physiological range.

We therefore define this mixture of weak acid and its conjugate base as a **buffer** because it is able to resist the large changes in pH that would have been observed if the same amount of base had been added to a solution of a strong acid at the same pH. For example, we see in this case that 400 ml of 0.1 M KOH–that is, 40 mmoles KOH–were required to raise the pH from 4 to 6. By contrast, less than 0.05 mmoles KOH would have been required to raise a comparable solution of strong acid from pH 4 to pH 6. Result: In the presence of buffer, almost 1000 times as much base would have been required to effect the same pH change.

> The concept of a buffer has been introduced by considering the titration of a weak acid with a strong base. We could equally well have constructed the titration curve by starting with a basic solution of the salt of the acid, i.e., 0.1 M K^+A^-, and titrating with a strong acid such as HCl. We would find that the pH drops sharply upon first addition of acid, enters a plateau phase, then drops sharply again once the neutralization point of the base is reached.

The key point is that in the plateau region of the titration curve, the buffer resists pH changes by either the addition of acid or base.

X. Properties of buffers

A. The common names for buffers. For convenience biochemists refer to buffers by the name of the conjugate base. (These names are shorter than those for conjugate acids.) For example, a buffer made from a mixture of acetic acid and potassium acetate is called "acetate buffer." A buffer made from imidazole (an amine) and imidazolium hydrochloride is called "imidazole buffer."

B. Meaning of the term "buffer concentration." When we say "0.1 M buffer" we mean that the *total* concentration of all species equals 0.1 M. For example, in a 0.1 M acetate buffer:
$[HAc]+[Ac^-] = 0.1$ M

C. Importance of the midpoint of the titration curve.

1. The midpoint of the titration, i.e., where $HA = A^-$, occurs at the pH that has the same numerical value as the pK_a. The reason for this is evident from the H-H equation: When HA = A, the ratio $(A^-/HA) = 1$, therefore:

$$pH = pK_a + \log 1$$

Since $\log 1 = 0$,

$$pH = pK_a$$

2. The middle of the curve is also the point where the buffering effect is strongest, i.e., where the change in pH upon addition of acid or base on pH is smallest.

Why is this so? Explanation:

The H-H equation shows that the pH of a solution of weak acid or base is determined not by the absolute concentration of conjugate acid and/or conjugate base, but by their *ratio*. A simple arithmetic example shows that this ratio changes maximally at the ends of the titration and minimally in the middle:

Consider a 0.1 M solution of a hypothetical acid HA, with $pK_a = 5$. The initial pH is approximately 3 and in one liter there are 99 mmoles of HA and 1mmol A^- at the start of the titration, so the starting ratio of A^-/HA is 1/99.

Now add KOH, 1mmole at a time. How will this affect the ratio? Remember that each addition converts 1mmole HA to 1mmole A^-, so the first addition changes the ratio from 1/99 to 2/98–nearly a two-fold increase! The next few additions make substantial but successively smaller fractional changes in the ratio: 3/97, 4/96, 5/95... The smallest incremental change occurs when the ratio approaches unity: 47/53, 48/52, 49/51, 50/50. As you can see, the incremental change here at the midpoint is only about 2%. Addition of more base this point causes the incremental change to rise in a mirror image of the first half of the titration curve: 51/49, 52/48, 53/47. As we reach the neutralization point, the incremental changes become as large as they were at the beginning: 95/5, 96/4, 97/3, 98/2, 99/1...

D. The end point in the titration of a weak acid or base is rarely pH 7 for several reasons, some of which are obvious and some of which are not. For convenience, we define the end point as the pH at which only 1% of the original species remains. This point occurs when the pH is 2 units above or below the pK_a. For example, if an acid has a pK_a of 4, the pH at which $(A^-/HA) = 100$ is 6, as shown by the H-H equation:

$$pH = 4+\log (100) = 4+2 = 6$$

E. Buffering capacity. The ability of a buffer to resist changes in pH is called the **buffering capacity.** There is no unique definition of it, but here is a good one: the number of moles of H^+ of OH^- that must be added to one liter of buffer in order to decrease or increase the pH by 1 unit.

Buffering capacity depends on two things: a) the ratio of conjugate base to conjugate acid, as was discussed in section C above, and b) the total concentration of species; a 1 M buffer has ten times the capacity to absorb H^+ of OH^- as does a 0.1 M buffer.

Because buffering capacity is highest at the pH equal to the pK_a, the buffer used for an experiment in the laboratory is one with a pK_a at or near the pH of the reaction being measured. This choice enables the experimenter to use the lowest concentration of buffer that will maintain pH. The "usable range" of a buffer is considered to be plus or minus 1 pH unit on either side of the pK_a.

F. Dilution of a buffer. According to the H-H equation, the pH of a buffer is dictated solely by the ratio of conjugate acid to conjugate base, and therefore one could dilute a buffer without changing the pH. In fact, large dilutions do change the pH somewhat. For our purposes, we will assume that modest changes in total buffer concentration (a few fold) do not change the pH significantly.

XI. Titration of Polyprotonic Acids and Bases

A. Many of the acids in biochemical systems have more than one ionizable proton. The pK_a's for each of these protons is different and, to the first order of approximation, each ionizing species may be considered as a separate acid that will dissociate completely before the next proton is removed. When the pK_a values differ by two or more units the error of this assumption is less than 1%.

Shown below is a titration curve for phosphoric acid:

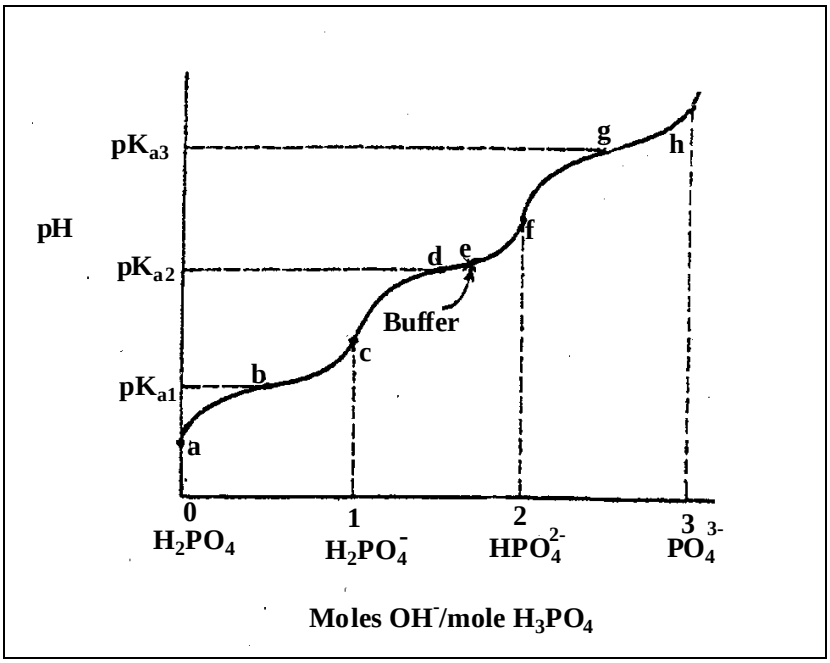


Figure 3: Titration curve of H_3PO_4.

At points a, c, f, and h, the predominant species are, respectively, phosphoric acid, monobasic phosphate, dibasic phosphate, and tribasic phosphate. These are the end-points of the titration curve. Points b, d, and g indicate the midpoints of three titrations, and lie at pH's corresponding to the pK_a's of the ionizations, i.e., at pH 2.1, 6.7, and 12.3. The most commonly used phosphate buffer is one slightly above pH 7, indicated by point e. This buffer is prepared from a mixture of the dibasic and monobasic forms of phosphate.

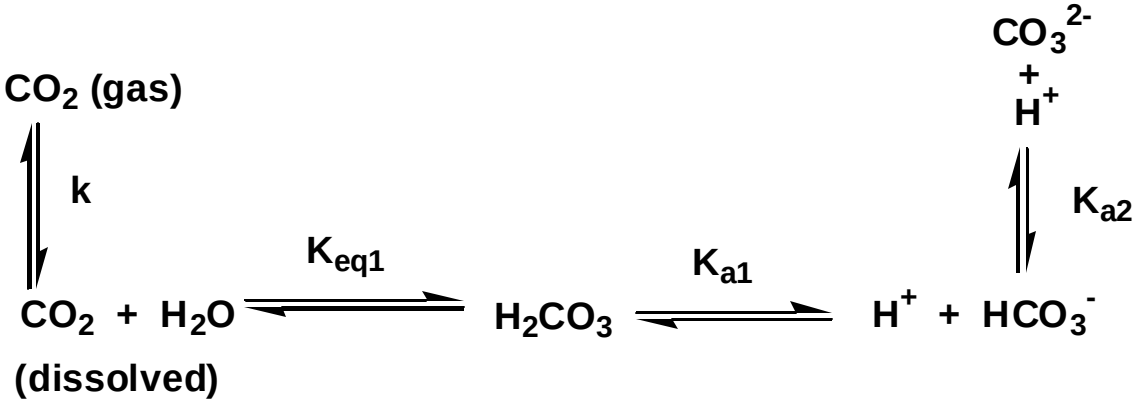
B. Blood Buffers--The HCO_3^-/CO_2 System

The HCO_3^-/CO_2 system is one of the two major blood buffers. Carbonic acid ionizes as a typical weak diprotic acid:

$$H_2CO_3 \underset{pK_{a1} = 3.8}{\overset{K_{a1} = 1.58 \times 10^{-4}}{\rightleftharpoons}} H^+ + HCO_3^-$$

$$HCO_3^- \underset{K_{a2} = 5.6 \times 10^{-11}}{\overset{pK_{a2} = 10.25}{\rightleftharpoons}} H^+ + CO_3^{2-}$$

However, most of the conjugate acid dissolved in blood and cytoplasm is present as CO_2, not HCO_3^-. The dissolved CO_2 is in equilibrium with CO_2 in the gas phase. A more complete presentation of the CO_2 buffer system is shown below.



The equilibrium between CO_2 (gas) and CO_2 (dissolved) is given by:

$$[CO_2]_{dissolved} = k(Pco_2)$$

That is, the concentration of dissolved CO_2 is directly proportional to the partial pressure of CO_2 in the gas phase. At 37°C and an ionic strength of 0.15, $k=3.01 \times 10^{-3}$ when Pco_2 is expressed in terms of mm Hg. The equilibrium constant for the reaction: dissolved $CO_2 + H_2O \rightleftharpoons H_2CO_2$ is about 5×10^{-3}:

$$K_{eq1} = \frac{[H_2CO_3]}{[CO_2]_{dis.}} = 5 \times 10^{-3}$$

Thus, the overall equilibrium constant between dissolved CO_2 and $H^+ + HCO_3^-$ is given by:

$$K'_a = \frac{[H^+][HCO_3^-]}{[CO_2]} = K_{eq1} \times K_{a1}$$

$$= (5 \times 10^{-3})(1.58 \times 10^{-4}) = 7.9 \times 10^{-7}$$

and $pK'_a = 6.1$

The relationship can also be written as:

$$K'_a = \frac{[H^+][HCO_3^-]}{(3.01 \times 10^{-5})P_{CO2}} = 7.9 \times 10^{-7}$$

At any pH:

$$pH = 6.1 + \log \frac{[HCO_3^-]}{[CO2]} \qquad \text{and} \qquad pH = 6.1 + \log \frac{[HCO_3^-]}{(3.01 \times 10^{-5})P_{CO2}}$$

For all practical purposes, a bicarbonate buffer can be considered to be composed of HCO_3^- (conjugate base) and dissolved CO_2 (conjugate acid).

The pH of blood is maintained at about 7.4. If the pK_a of CO_2 is 6.1, how can the HCO_3^-/CO_2 help buffer blood at pH 7.4? Everything we have learned so far suggests that a buffer is effective only in the region of its pK_a. The key here is that *in vivo* the HCO_3^-/CO_2 buffer is an *open system* in which the concentration of dissolved CO_2 is maintained constant. Any excess CO_2 produced by the reaction $H^+ + HCO_3^- \rightarrow H_2O + CO_2$ is expelled by the lungs. In contrast, the usual laboratory buffer is a closed system. The concentration of conjugate acid increases when H^+ reacts with the conjugate base. The effectiveness of the open system is illustrated below.

Problem 1:

Blood plasma contains a total carbonate pool (essentially $HCO_3^- + CO_2$) of 2.52×10^{-2} M. (a) What is the HCO_3^-/CO_2 ratio and the concentration of each buffer component present at pH 7.4? (b) What would the pH be if 10^{-2} M H^+ is added under conditions where the increased $[CO_2]$ cannot be released? (c) What would the pH be if 10^{-2} M H^+ is added and the excess CO_2 eliminated (thereby maintaining the original $[CO_2]$)?

Solution:

(a)

$$pH = pK'_a + \log \frac{[HCO_3^-]}{[CO_2]} \qquad 7.4 = 6.1 + \log \frac{[HCO_3^-]}{[CO_2]}$$

$$1.3 = \log \frac{[HCO_3^-]}{[CO_2]} \qquad \boxed{\frac{[HCO_3^-]}{[CO_2]} = \frac{20}{1}}$$

$$[HCO_3^-] = \frac{20}{21} \times 2.52 \times 10^{-2} = \boxed{2.40 \times 10^{-2}\ M}$$

$$[CO_2] = \boxed{1.2 \times 10^{-3}\ M}$$

(b) If 0.01 M H^+ is added:

$$[HCO3^-]_{final} = 0.024 - 0.010 = 0.014 \text{ M}$$

$$[CO_2]_{final} = 0.0012 + 0.010 = 0.0112 \text{ M}$$

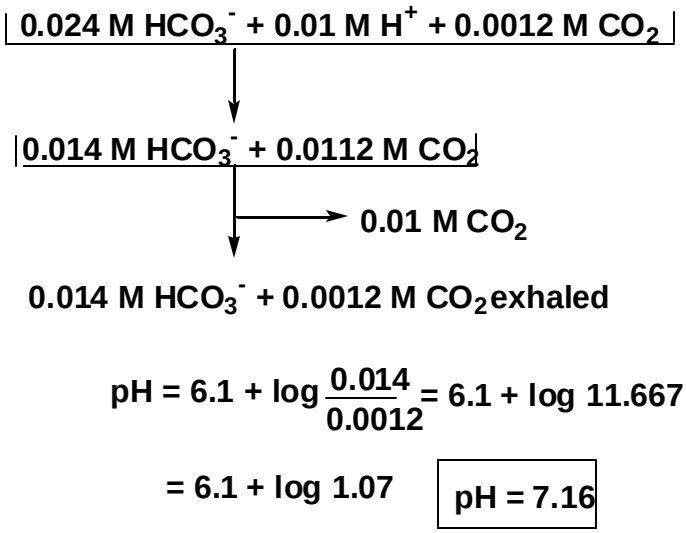
$$pH = 6.1 + \log \frac{0.014}{0.0012} = 6.1 + \log 1.25$$

$$= 6.1 + 0.097$$

$$\boxed{pH = 7.16}$$

Clearly, in a closed system, the HCO_3^-/CO_2 mixture has very little buffer capacity at pH 7.4.

(c) In an open system:



$$pH = 6.1 + \log \frac{0.014}{0.0012} = 6.1 + \log 11.667$$

$$= 6.1 + \log 1.07 \qquad \boxed{pH = 7.16}$$

In an open system, the pH decreases only 0.24 pH unit. At first glance, it would seem that in an open system, the HCO_3^- reserve would be rapidly depleted. However, *in vivo,* HCO_3^- is constantly replenished by the oxidative metabolic pathways.

XII. Problems, Exercises:

A. Define and give examples where appropriate

1. Clinical acidosis/alkalosis
2. Molarity
3. Bronsted acid and base
4. Conjugate acid-conjugate base pair
5. Strong acid and strong base
6. Weak acid and weak base
7. The ion product of water
8. Neutrality
9. pH
10. pH scale
11. Titration, titration curve
12. Monoprotic acid
13. Polyprotic acid
14. pK_a
15. Henderson-Hasselbach Equation
16. Buffer
17. The term "buffer concentration"
18. Buffering capacity
19. End point/equivalence point
20. Neutralization

B. Explain

1. Why do buffers, which are made from weak acids and bases, resist pH changes better than strong acids and bases in the physiological range of pH?

2. A series of enzyme assays is to be performed at pH 7.0. Why would phosphate be a good choice of buffer? Which species of phosphate would predominate at pH7?

C. Calculate:

1. The pH for hydrogen ion concentrations of

   a. 1×10^{-6} M
   b. 2×10^{-6} M
   c. 4×10^{-6} M
   d. 6×10^{-6} M
   e. 8×10^{-6} M
   f. 1×10^{-5} M

2. The hydrogen ion concentration for the normal limits of blood pH: 7.37 and 7.43.

3. The hydrogen ion concentration at $[OH^-] = 5 \times 10^{-5}$ M.

4. The approximate pH of a 10^{-8} M solution of HCl (this is tricky).

5. The endpoint for the titration of NH_3 with a strong acid.

6. The ratio of acetate ion/acetic acid in an acetate buffer at pH 3, 4, 5, and 6.

7. The final pH when 5 mmoles KOH is added to 1 liter of 4×10^{-3} M HCl. (Assume for simplicity that the final volume is not changed significantly.)

Answers:

A. Definitions can be found in bold type in the text of the handout.

B1. The $[H^+]$ in a solution of strong acid or strong base is directly proportional to amount of acid or base added. If the concentration is low, as it is in the physiological range, addition of small amounts of acid or base makes a large relative change.

In contrast, the pH of a buffer is determined by the ratio of conjugate acid to conjugate base. Both species are usually present in relatively high concentration. Under these conditions, adding small amounts of acid or base makes a relatively small change in the ratio of conjugate acid to conjugate base, and therefore makes a relatively small change in the pH.

B2. The $pK_{a(2)}$ for phosphate is 6.7, which is within one log unit of the desired pH. Phosphate should therefore provide adequate buffering. The predominant species are the monobasic and dibasic forms.

C1. $pH = -\log [H^+]$. Therefore, the pH values for the six solutions are: a) 6; b) 5.7; c) 5.4; d) 5.22; e) 5.1; f) 5.0.

C2. $[H^+]$ = antilog (-pH). Therefore, the $[H^+]$ at pH 7.37 + 4.26 X 10^{-8}, and the $[H^+]$ at pH 7.43 = 3.71 x 10^{-8}.

C3. $K_w = [H^+][OH^-] = 10^{-14}$. Therefore, the $[H^+] = 10^{-14}/5 \times 10^{-5} = 2 \times 10^{-10}$.

C4. The answer is not pH = 8. This answer would be correct if pure water contained zero H^+. But pure water contains 10^{-7} M hydrogen ions. Therefore, the pH would be slightly less than 7. As a rough approximation we could simply add the contribution of protons form water and from the HCl: $10^{-7} + 10^{-8} = 1.1 \text{ X } 10^{-7}$; pH = 6.95.

C5. For convenience we define the endpoint as two log units on either side of the pK_a. Since the pK_a for ammonia (to be precise, the ammonium ion) is 9.4, the acid end point is 7.4.

C6. Use the H-H equation:

(Acetate ion/acetic acid) = antilog ($pH-pK_a$)

The pK_a for acetic acid is 4.74. Therefore, (Ac^-/HAC) at pH 3, 4, 5, and 6 increases as follows: 0.018, 0.18, 1.8, 18. Notice how the ratio increases by a factor of 10 for each rise in the pH by one unit.

C7. There are 4 mmoles HCl at the start. Adding 5 mmoles KOH produces a solution with an excess of 1 mmole KOH. If the volume is still approximately 1 liter, the $[OH^-] = 1 \times 10^{-3}$ M.

The $[H^+] = 10^{-14}/10^{-3} = 10^{-11}$, corresponding to pH 11.

A BRIEF REVIEW OF COMMON LOGS

Definitions

log

log means "logarithm to the base 10." Thus, log y = x means $y = 10^x$. In plain language, we can understand the function this way:

Question: What is the log of y?
Answer: It is the power of 10, in this case x, that produces y.

So, for example, log 100 = 2 because 2 is the power of 10 that produces 100 (i.e., 10^2). More examples (make sure you understand these):

log 1000 = 3
log 100 = 2
log 10 = 1
log 1 = 0
log 0.1 = -1
log 0.01 = -2
log 0.001 = -3

You cannot have the log of a negative number, or the log of zero. See if you can figure out why, using numerical examples.

antilog

The antilog is the inverse function of log. Thus, antilog (log y) = y. In similar fashion, log (antilog x) = x. The definition of antilog x is also 10^x, as we can demonstrate below.

Consider the numerical example:

log 100 = 2

Take the antilog of each side of the equation:

antilog (log 100) = antilog 2

100 = antilog 2

For this equation to hold, antilog 2 must be 10^2. Thus, antilog x = 10^x. Note that the antilog button on your calculator is 10^x.

Rules of logs

Rule 1: $\log y^n = n \log y$

Examples: $\log 2^3 = 3 \log 2 = 3 (0.693) = 1.386$

$\log 10^6 = 6 \log 10 = 6$

$\log 10^{-7} = -7 \log 10 = -7$

Rule 2: $-\log x = \log 1/x$

This can be easily demonstrated by application of the law of exponents and of Rule 1:

$\log 1/x = \log x^{-1}$

$(-1)(\log x) = - \log x$

Common application: If the $[H^+] = 10^{-6}$, what is the pH?

Answer: $pH = -\log [H^+] = \log (1/[H^+]) = \log 6 = 6$

Rule 3: $\log ab = \log a + \log b$

Numerical example: $\log (100)(1000) = \log 100{,}000 = 5 = \log 100 + \log 1000$

Rule 4: $\log a/b = \log a - \log b$

Numerical example: $\log 1000/100 = \log 10 = 1 = \log 1000 - \log 100$

Ionization of Amino Acids & Peptides

Ionization of Amino Acids and Peptides

(Recommended Reading: Lehninger pp 123-126)

(See also: http://cti.itc.virginia.edu/~cmg/Demo/markPka/markPkaApplet.html)

I. Introduction

The common amino acids have 2-3 ionizable groups, and therefore exist as a mixture of ionic forms. For at least two reasons, it is important to understand how the distribution of ionic forms varies with pH and pK_a. First, as stated earlier, the charge on amino acids affects the structure and activity of proteins. Second, many of the techniques used to identify and seperate proteins are based upon differences in charge. One of these techniques is ***electrophoresis.*** In a typical application, a small volume of sample is put into a well in a thin slab of gel coating a glass plate. Electrodes are placed at either end of the plate, creating a potential gradient. Proteins that bear a net negative charge migrate towards the positive electrode (anode), whereas those that bear a positive charge migrate towards the negative electrode (cathode).

There is always a pH called the ***pI*** at which the sum of the negative charges on an amino acid, peptide, or protein exactly balances the sum of the positive charges. When the net charge on the molecule is zero, it does not migrate in an electric field and is said to be at its ***isoelectric point***.

II. Exercises

This hour will be essentially a problem solving session. We will work through the titration curves of several amino acids, and figure out what the predominant ionic forms are at key points on the curves. Eventually you should be able to work backwards: sketch the titration curve for any amino acid from a table of pK_a's. We will also go through the methods for calculating and estimating the pI of any amino acid or peptide.

Table 1 (5-1 in Lehninger) has accurate pK_a data for practicing certain homework problems with a calculator. Table 2 in the notes contains approximate data for working problems without a calculator. Many students have a harder time with problems involving estimation (which happen to be the ones that show up on an exam) than those requiring detailed calculation. Sampler:

> A tripeptide has a net charge of - 0.60 at pH 7.0. The pI must be
>
> a. > 7.0
> b. < 7.0
> c. Not enough information to decide between a or b.
>
> Justify your answer.

***

Table 1: Properties and Conventions Associated with the Standard Amino Acid

Amino Acid	Abbreviated names	M_r	pK_a values: pK_1 (-COOH)	pK_2 ($-NH_{3+}$)	pK_R (R group)	pI	Hydropathy index*	Occurrence in proteins (%)†
Nonpolar, aliphatic R groups								
Glycine	Gly G	75	2.34	9.6		5.97	-0.4	7.2
Alanine	Ala A	89	2.34	9.69		6.01	1.8	7.8
Valine	Val V	117	2.32	9.62		5.97	4.2	6.6
Leucine	Leu L	131	2.36	9.6		5.98	3.8	9.1
Isoleucine	Ile I	131	2.36	9.68		6.02	4.5	5.3
Methionine	Met M	149	2.28	9.21		5.74	1.9	2.3
Aromatic R groups								
Phenylalanine	Phe F	165	1.83	9.13		5.48	2.8	3.9
Tyrosine	Tyr Y	181	2.2	9.11	10.07	5.66	-1.3	3.2
Tryptophan	Trp W	204	2.38	9.39		5.89	-0.9	1.4
Polar, uncharged R groups								
Serine	Ser S	105	2.21	9.15		5.68	-0.8	6.8
Proline	Pro P	115	1.99	10.96		6.48	1.6	5.2
Threonine	Thr T	119	2.11	9.62		5.87	-0.7	5.9
Cysteine	Cys C	121	1.96	10.28	8.18	5.07	2.5	1.9
Asparagine	Asn N	132	2.02	8.8		5.41	-3.5	4.3
Glutamine	Gln Q	146	2.17	9.13		5.65	-3.5	4.2
Positively charged R groups								
Lysine	Lys K	146	2.18	8.95	10.53	9.74	-3.9	5.9
Histidine	His H	155	1.82	9.17	6	7.59	-3.2	2.3
Arginine	Arg R	174	2.17	9.04	12.48	10.76	-4.5	5.1
Negatively charged R groups								
Aspartate	Asp D	133	1.88	9.6	3.65	2.77	-3.5	5.3
Glutamate	Glu E	147	2.19	9.67	4.25	3.22	-3.5	6.3

*A scale combining hydrophobicity and hydrophilicity of R groups; it can be used to measure the tendency of an amino acid to seek an aqueous environment (- values) or a hydrophobic environment (+ values). See Chapter 12. From Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol*. **157**, 105-132. †Average occurrence in over 1150 proteins. From Doolittle, R.F. (1989) Redundancies in protein sequences. In *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G.D., ed) Plenum Press, NY, pp. 599-623.

Table 2: Approximate pK_a's of Amino Acids

Amino Acid	pK_{a1} (Carboxyl)	pK_{a2} (Amino)	pK_{a2} (R-group)
Nonploar aliphatic	2.0	9.5	----
Glycine			
Proline			
Alanine			
Leucine			
Valine			
Isoleucine			
Aromatic	2.0	9.5	---
Phenylalanine			
Tyrosine			10.0
Tryptophan			
Polar, uncharged	2.0	9.5	—
Cysteine			8.0
Methionine			
Threonine			
Serine			
Asparagine			
Glutamine			
Charged-negative	2.0	9.5	4.0
Aspartate			
Glutamate			
Charged-positive	2.0	9.5	—
Histidine			6.0
Lysine			10.5
Arginine			12.5

A. Titration Curves. We will discuss these in class.

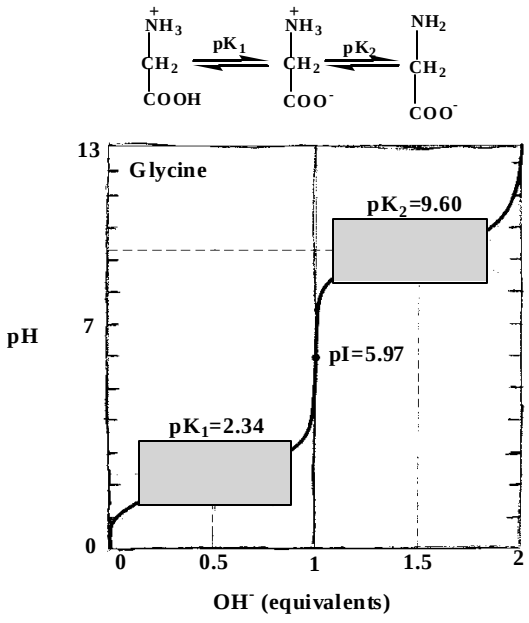


Figure 1: Titration of an amino acid. Shown here is the titration curve of 0.1 M glycine at 25°C. The ionic species predominating at key points in the titration are shown above the graph. The boxes, centered at about pK_1 = 2.34 and pK_2 = 9.60, indicating the regions of greatest buffering power.

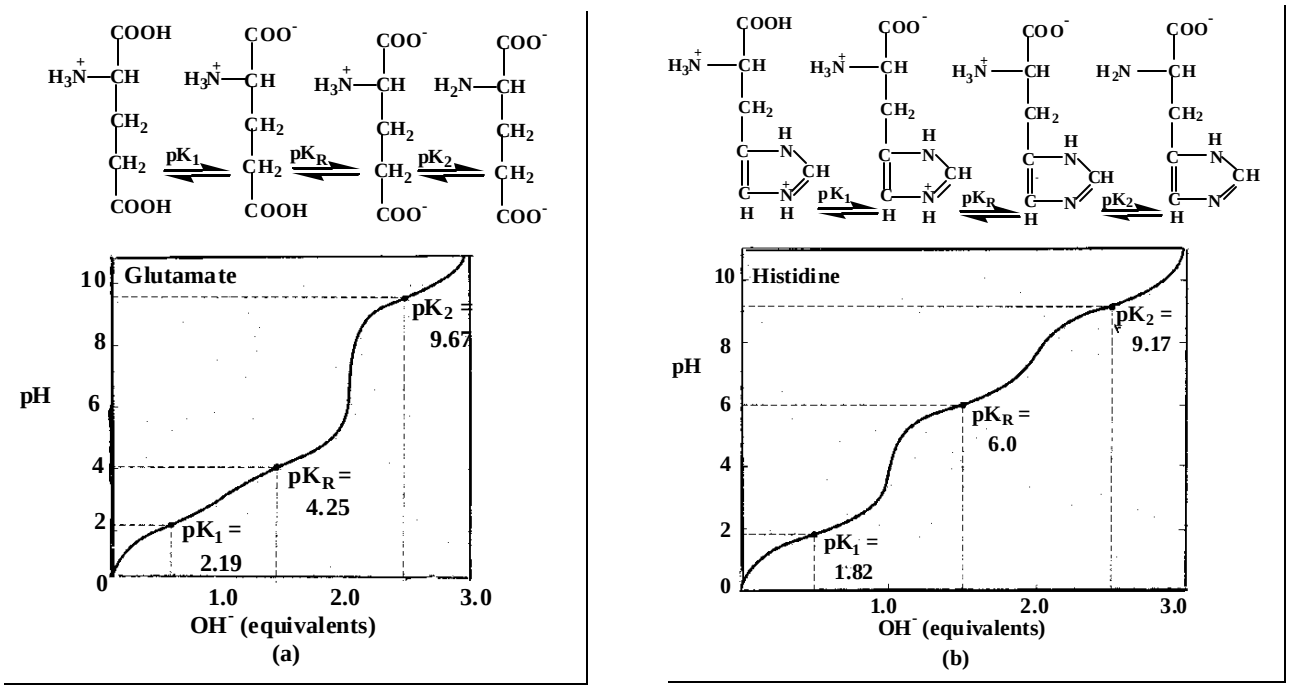


Figure 2: Titration curves for (a) glutamate and (b) histidine. The pK_a of the R group is designated here as pK_R.

B. Calculating pI's

On an exam, you will be required to *estimate* pI's of amino acids and peptides, not calculate them exactly. (Unless the arithmetic is easy.) But to understand the estimation procedure, you may find it helpful to see how a rigorous calculation is done. Therefore, we will discuss the calculation method first.

Overview

The basic idea is to start by selecting a test pH, such as 7.0. Then calculate the **fractional charge** of each functional group of the molecule (see below). Add the fractional charges together to get the *net* charge on the molecule. If the sum is negative, try a lower pH (Why?). If the sum is positive try a higher pH. By trial and error, you can find pI within a minute, if you program this operation into a spreadsheet.

Details

How to calculate fractional charge:

The fractional charge of a functional group, N, tells you the sign and the fraction of the population in the charged form. For example, if a snapshot of carboxyl groups shows 80 out of 100 are in the ionized form, then the fractional charge for the carboxyl is - 0.8. What this also means physically (owing to dynamic equilibrium) is that each carboxyl spends 80% of its time ionized and 20% of its time unionized, so each carboxyl migrates in an electric field as if it had a charge of - 0.8.

From this numerical example, we see that N for the carboxyl group is defined mathematically as

$$N = [A^-]/([AH] + [A^-]) = [A^-]/[A_t] \qquad \text{Eqn 1}$$

where A_t is the sum of all ionic forms. This formula also applies to all functional groups that exist as a mixture of neutral and anionic forms.

Similarly, for an amino group or any other group that exists as a mixture of neutral and cationic forms:

$$N = [BH^+]/([BH^+] + [B]) = [BH^+]/[B_t] \qquad \text{Eqn 2}$$

The *net* charge on an amino acid, peptide, or protein, is the sum of the fractional charges.

Sample problem:

Calculate the net charge of aspartic acid at pH = 3.50, using the pK_a values from Lehninger: $pK_{a1} = 1.88$; $pK_{a2} = 9.60$; $pK_{aR} = 3.65$.

(1) Calculate N for the "- carboxyl

The H-H equation gives us the ratio, $[A^-]/[HA]$:

$pH = pK_a + \log [A^-]/[HA]$

$[A^-]/[HA]$ = antilog (3.50 - 1.88)
= 41.7

Since 41.7 can be expressed as 41.7/1 , i.e., 41.7 A^-: 1 HA, the fraction of molecules in the A^- form is 41.7/(41.7 +1). Therefore,

N = - 0.976.

(2) Calculate N for the side chain carboxyl

$[A^-]/[HA]$ = antilog (3.50 - 3.65)
= 0.708

$[A^-]/[A_t] = 0.708/(1 + 0.708)$

N = - 0.414

(3) Calculate N for the "- amino group

$pH = pK_a + \log [BH]/[B^+]$

$[BH]/[B^+]$ = antilog (3.50 - 9.60)

$= 7.94 \times 10^{-7}$

We need the inverse of this ratio for our calculation of N (Eqn 2) :

$[B^+]/[BH] = 1.26 \times 10^6$

$[B^+]/[B_t] = 1.26 \times 10^6/(1 + 1.26 \times 10^6)$

N . + 1

(4) Sum the fractional charges

net charge = - 0.976 - 0.414 + 1

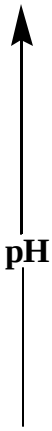
= - 0.390

We find that the charge is not far from zero. As it turns out, the exact pI is midway between the pK_a's of the carboxyl groups: (1.88 + 3.65)/2 = 2.76. At this pH, the sum of the fractional charges of the carboxyls is almost exactly -1, offsetting the +1 charge of the amino group. See if you can demonstrate this by calculation.

C. Estimating pI

The approach here is the same as above, except that approximate N values are used, such as those in Table 3:

Table 3



pH-pK_a	N_As	N_{BH}+	
> +1	-1	0	
+1	-0.9	+0.10	
+0.5	-0.75	+0.25	⇐ pK_a
0	-0.5	+0.5	
-0.5	-0.25	+0.75	
-1.0	-0.10	+0.9	
> -1.0	0	+1	

Procedure. Make pH 7.0 your first test pH and compute the net charge. If it is negative, try a lower pH; if it is positive, try a higher pH. Iterate until you find the range of the pI.

Sample problem: What is the approximate pI of the peptide asp-his ?

(1) Remember that a peptide always has a terminal amino group and terminal carboxyl group, i.e., NH_2 -asp-his- COOH. The internal "- amino and carboxyl groups are joined in an amide linkage, and don't ionize. Assume the pK_a's for the terminal amino and carboxyl are 9.5 and 2.0.

There are 4 ionizable groups in this molecule: the N-terminal amino, the C-terminal carboxyl, and the R groups.

(2) Tabulate the functional groups as follows:

Table 4

Functional group	pK_a	Test pH	$pH-pK_{a1}$	N
C-term	2.5	7.0	> 1.0	-1
R_1(asp)	4.5		>1.0	-1
R_2 (his)	6.0		1.0	+0.1
N-term	9.5		> -1.0	+1
Net				**~ -1**

We find that the peptide has a net charge of about -1 at pH 7.0. Therefore, the pI must be lower. How much lower? What would be a reasonable choice for the next test pH? pH 5.0 would be a good one. The logic: We have to eliminate a single negative charge–the equivalent of gaining a positively charged group or eliminating a negatively charged group. We can gain a positively charged group by going well below the pK_a of 6 for his. But we don't want to go too low in pH or we will start neutralizing the charge on the side change of asp.

Table 5

Functional group	pK_a	Test pH	$pH-pK_{a1}$	N
C-term		**5.0**	**> 1.0**	**-1**
R_1(asp)			**0.5**	**-0.75**
R_2 (his)			**1.0**	**+0.9**
N-term			**> 1.0**	**+1.0**
Net				**+0.15**

The net charge has changed sign, and is rather small–only +0.15. Because the net charge has changed sign, we know the pI must be between 5 and 7. What would be the best choice for the next test pH: 5.5? 6.0? 6.5? To be discussed in class.

Practice problems:

1. The pI's for amino acids without ionizable side chains are exactly midway between the pKa's of the "- amino and "- carboxyl groups:

   $(pK_{a1} + pK_{a2})/2$.

   Why do the fractional charges of the two functional groups exactly balance at this point? Use the H-H equation to explain. [Hint: What is $(pH - pK_a)$ for each group?]

2. Suggest simple formulas for calculating exactly the pI's of his, arg, lys, tyr, and cys. Explain.

3. What are the approximate pI's for the following peptides, to the nearest 0.5 pH unit? All of these can be worked by estimation.

   a. ala-lys
   b. ala-lys-lys.
   c. glu-his
   d. ala-arg-asp-asp-asp-asp

Answers to practice problems:

1. At the midpoint between the pK_a's of the "-amino and "-carboxyl groups, $(pH - pK_a)_{COOH} = (pK_a - pH)_{NH2}$. Therefore, from the H-H equation, log $[A^-]/[HA]$ = log $[BH^+]/[B]$. Since the fractional charge is determined by these ratios, $N_{COOH} = N_{NH2}$.

2. As pointed out earlier, a simple formula gives the exact pI for aspartic acid. It also applies to glutamic acid:

$$(pK_{"\text{-COOH}} + pK_{\text{side chain COOH}})/2.$$

The reason is that midway between the pK_as of the carboxyl groups, the sum of the fractional charges is exactly - 1.00 (You can demonstrate this with numerical examples), which balances the full positive charge on the " amino group.

A similar logic applies to the formulas for his, arg, and lys, except in these cases you need to find a pH where the sum of the positive fractional charges is +1, balancing the -1 charge on the "-carboxyl:

$$(pK_{"\text{-NH2}} + pK_{\text{side chain BH+}})/2.$$

The formulas for tyr and cys are similar to those for asp and glu:

$$(pK_{"\text{-COOH}} + pK_{\text{side chain AH}})/2.$$

3. Estimate the pI's of

a. ala-lys

You can always work out a problem like this using the method outlined in the sample problem. But here is a super short-cut: When you have just 2 or 3 functional groups, use the same procedure as you would for the free amino acids. In this instance, the molecule has two functional groups that ionize to form cations (N-terminus and lysine side chain) and one that forms an anion (C-terminus). So the pI must be midway between the pKs of the cation formers: (9.5 + 10.5)/2 = 10.

b. ala-lys-lys

There are 4 functional groups, one of them negative at pH 7, and the others positive:

C-term, N = -1
N-term, N = +1
lys side chain N = +1
lys side chain N = +1
lys side chain N = +1

The net charge is +3. So the pI must be >7, and high enough to eliminate 3 positive charges. As you increase the pH, the first pK_a you encounter is that for the N-terminal amino, at 9.5. At this pH you eliminate 0.5 positive charge from the terminal amino, but have not yet significantly removed protons from the lysine side chains (pK_a = 10.5). Keep increasing the pH. Try pH =10.5:

C-term, N = -1
N-term, N = + 0.1
Lys N = + 0.5
Lys N = + 0.5
Lys N = + 0.5

Net charge + 0.6. Getting close. What about pH 11? Obviously too high because there would remain only + 0.1 charge on each of the lys side chains, so the net charge would become negative. Therefore, the pI must be between 10.5 and 11.

c. glu-his

At pH 7:

C-term, N = -1
glu side chain N = -1
his side chain N = + 0.1
N-term N = + 1

Net charge = - 0.9

The pI must be < 7, and far enough below 7 to eliminate 0.9 negative charge. As we lower the pH, the his is titrated first, with a pK_a = 6. We have to eliminate almost a full charge, so try pH 5.0:

C-term N = -1
glu N = -1
his N = + 0.9
N-term N = + 1

Net charge, -0.1.

Still negative, if we go to pH 4.5, the net charge must change sign because his increases to +1 and glu decreases to -0.9. Therefore, the pI is between 4.5 and 5.0.

d. ala-arg-asp-asp-asp-asp

At pH 7

C-term	N = -1
asp	N = -1
asp	N = -1
asp	N = -1
asp	N = -1
arg	N = +1
N-term	N = +1

Net charge = -3

To eliminate the negative charge we have to reduce the pH. The first titrateable group is asp, with $pK_a = 4$, and the second is the C-term, with $pK_a = 2$. If we go too far below the pK_a of asp, we will eliminate 4 negative charges–one too many. The trick is to find a pH where the C-term contributes one negative charge, and each asp contributes -0.25. Answer: pH 3.5.

Noncovalent Forces Between Molecules & Properties of Water

NONCOVALENT FORCES BETWEENMOLECULES
AND PROPERTIES OF WATER

(Recommended Reading: Ch. 4, Lehninger)

I. Introduction

Proteins are polymers that twist and fold into a variety of three dimensional shapes: rods, sheets, barrels, and quasi-spherical globules to name a few. The forces that stabilize these structures are mainly non-covalent interactions, which are usually categorized as: electrostatic, van der Waals, hydrogen bond, and hydrophobic. These forces determine not only the three dimensional structures of proteins, but also the three dimensional structures of other macromolecules such as DNA. They are also largely responsible for holding together the components of supra molecular structures, such as biological membranes.

Perhaps the most common and important phenomenon in biochemistry and cell biology is ***binding***; substrates bind to enzymes, hormones bind to receptors, antibodies bind to antigens, and so on. In most cases, noncovalent forces are responsible for these interactions.

The interactions of molecules in a cell take place in a variety of media: water, membranes, fat droplets, and the interfaces between water and membranes or membrane-like structures. The relative importance of the various forces is different in each medium.

To understand what these forces are and how they function to create order in the cell, it is necessary to start from first principles and work our way up to more complex systems.

II. Description of the forces

The most obvious, everyday manifestation of intermolecular forces is the very existence of solids and liquids. If molecules did not attract each other, they would exist exclusively in a gaseous state.

Sodium chloride exists as a crystal because of an electrostatic attraction between the positively charged sodium ions and negatively charged chloride ions; water exists as a solid (ice) or a liquid mainly because of intermolecular hydrogen bonds between oxygen and hydrogen atoms; hydrocarbons can exist as either liquids or solids as a result of van der Waals interactions.

A. Electrostatic Interactions

Opposite charges attract and like charges repel. The energy of the interaction, U, is given by Coulomb's Law

$$U = \frac{kq_1q_2}{Dr}$$

where q_1 is the charge on one particle and q_2 is the charge on the other; r is the distance separating the particles; D is the dielectric constant of the medium; k is the a constant of proportionality.

The dielectric constant is a measure of the ability of the medium to screen the charges from one another. The smallest value of D is unity, found in a vacuum. A hydrocarbon solvent has a value of 2-3, whereas water has a value of 80. The higher the dielectric constant, the more effective is the solvent at screening charges, and therefore the weaker is the interaction of the charges with each other (notice that D is in the denominator of the equation). Consequently, charged species are much more soluble in solvents of high dielectric than in solvents of low dielectric.

All the functional groups in organic molecules (except C=C and C/C bonds) contain ***heteroatoms***--nitrogen, oxygen, sulfur--and are therefore polarized; the heteroatom attracts the shared electrons of the covalent bond more than the carbon does. For example, in the carbonyl group, the oxygen attracts the bonding electrons more than the carbonyl carbon; consequently, the oxygen carries a partial negative charge whereas the carbon carries a partial positive charge. Because of this charge separation, the carbonyl group has a permanent ***dipole***. Dipoles attract each other electrostatically if they are antiparallel, and repel each other if they are parallel. If they are free to rotate, the antiparallel arrangement will predominate and there will be a net attractive force between the molecules.

B. Hydrogen bonds and the structure of water

Hydrogen bonds have been mentioned briefly; we will discuss them in detail now.

A hydrogen bond is a special kind of electrostatic bond between two dipoles. In general, a hydrogen bond may be represented as D-H···A, where D-H is a weakly acidic "donor group" such as N-H or O-H, and A is an "acceptor" atom such as N or O, which bears a lone pair of electrons.

Shown below is a H-bond between two water molecules.

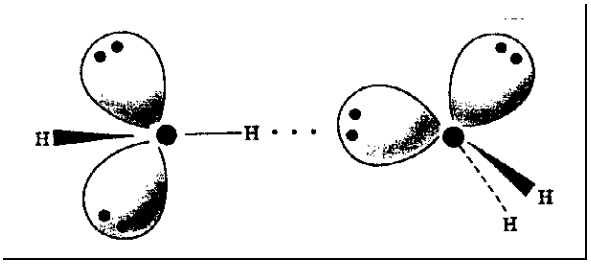


Figure 1: A hydrogen bond between two water molecules. The strength of this interaction is maximal when the O–H covalent bond points directly along a lone pair electron cloud of the oxygen atom to which it is hydrogen bonded.

Notice that the O-H bond of the donor points towards the oxygen atom of the acceptor along the lone pair electron cloud. This is the most stable geometry for a hydrogen bond, although weaker "bent" hydrogen bonds are also common. The bond is most stable when the distance between the H and the acceptor is about 1.8 Å. The fact that the H-bond has a preferred orientation and optimal length has profound implications for the structures of macromolecules and for the ***specificity*** of binding and interaction. Perhaps the most famous example is the pairing of complementary bases in DNA Remember the story of James Watson sitting with cardboard cutouts of the bases A, T, C, and G, trying to see which pairs would interact via H-bonds. What he discovered was that A matched with T, and C matched with G:

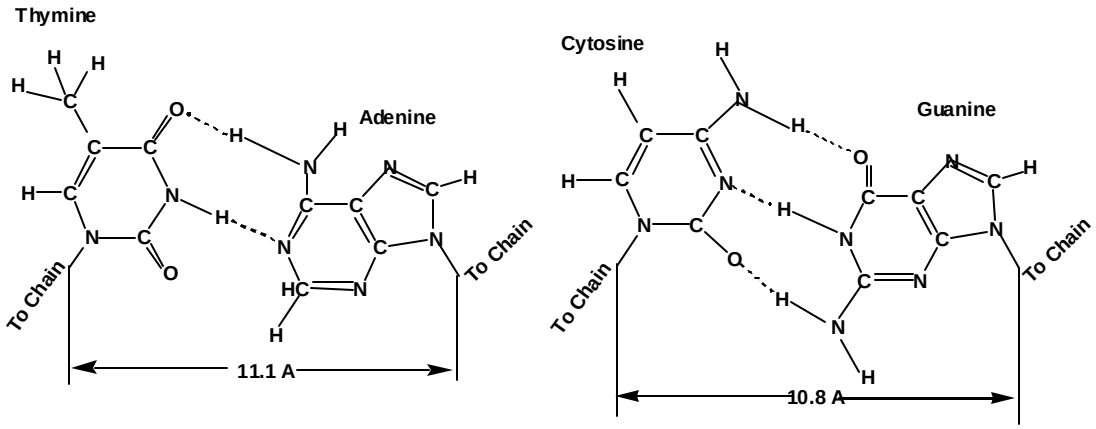


Figure 2: Base pairs of DNA. Note that the pyrimidine bases are "flipped over" from the positions in which they are usually shown. The bases must be in this orientation to form base pairs.

The C=O···H-N hydrogen bonds appear to be bent because the 4 atoms in the structure are not co-linear, but remember that linearity in a H-bond refers to line between the donor D-H bond and the lone-pair of electrons of the acceptor. In a carbonyl group, the lone pair electrons on oxygen sticks out at an angle to the C=O bond.

Back to water...

The two hydrogens of water can serve as donors, and the two sets of lone-pair electrons can serve as H bond acceptors. The donors and acceptors are arranged tetrahedrally. In ice, all the potential hydrogen bonds are formed, creating an open, regular lattice as shown below:

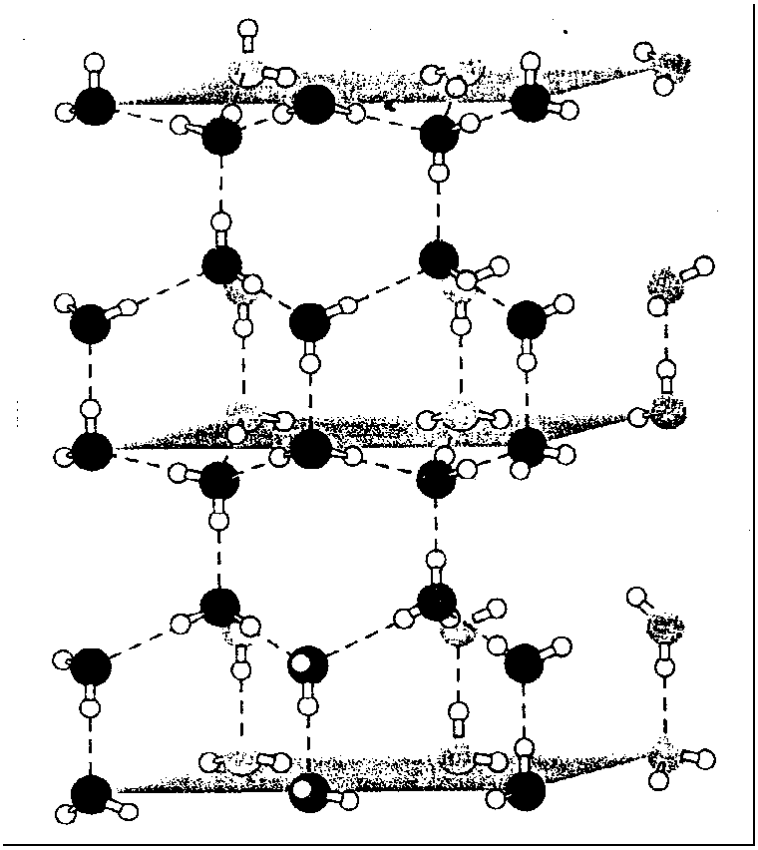


Figure 3: The structure of ice. The tetrahedral arrangement of the water molecules is a consequence of the roughly tetrahedral disposition of each oxygen atom's sp^3-hybridized bonding and lone pair orbitals. Oxygen and hydrogen atoms are represented, respectively, by black and white spheres, and hydrogen bonds are indicated by dashed lines. Note the open structure that gives ice its low density relative to liquid water. [After Pauling, L., *The Nature of the Chemical Bond* (3rd ed.), p. 465, Cornell University Press (1960).]

In the liquid state, water may be thought of as existing in "flickering clusters" of smaller hydrogen bonded structures. These form and break up almost as fast as a molecule can tumble--on the order of picoseconds. Nevertheless, liquid water at 0° contains 85% of the H-bonds of ice.

How important are H-bonds in maintaining water as a liquid? Without them, water would have approximately the same boiling point as methane: -183 °C. The high boiling point of water reflects its extraordinary internal cohesiveness resulting from intermolecular hydrogen bonding.

C. Van der Walls forces

The noncovalent associations between electrically neutral molecules are collectively known as ***van der Waals*** interactions. We have discussed two of these: dipole-dipole, and hydrogen bonds. Hydrogen bonds are usually considered as a separate category.

There is also a positive force of attraction between a permanent dipole and a nonpolar molecule, and a force of attraction between nonpolar molecules. These are illustrated below and explained in the figure legend.

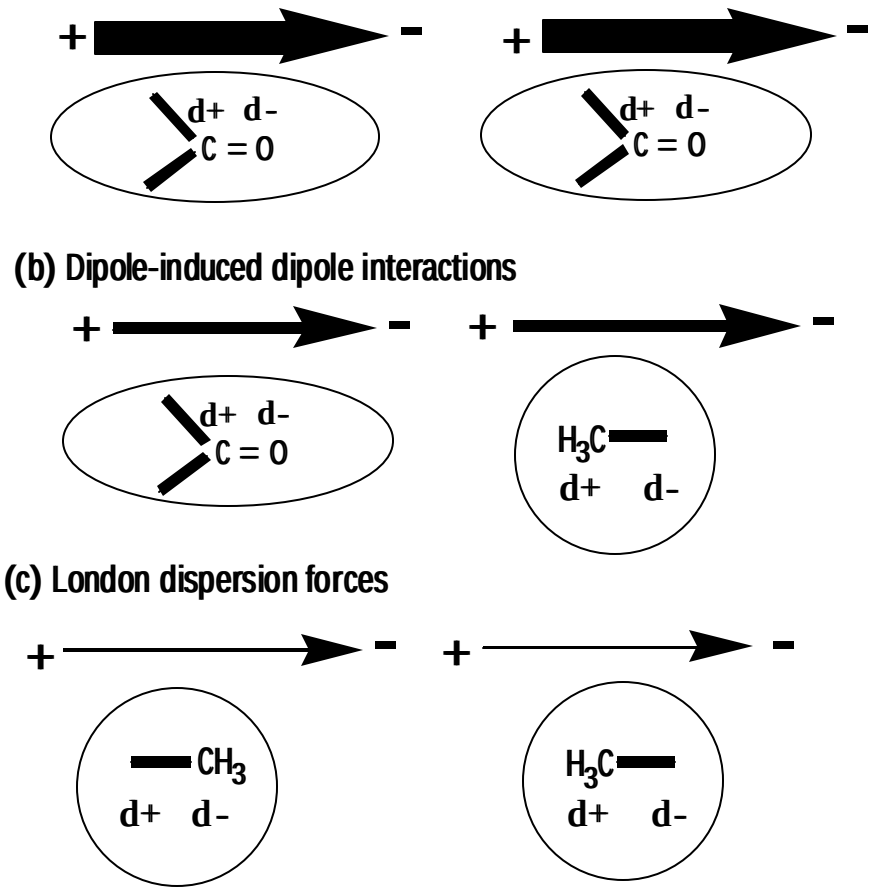


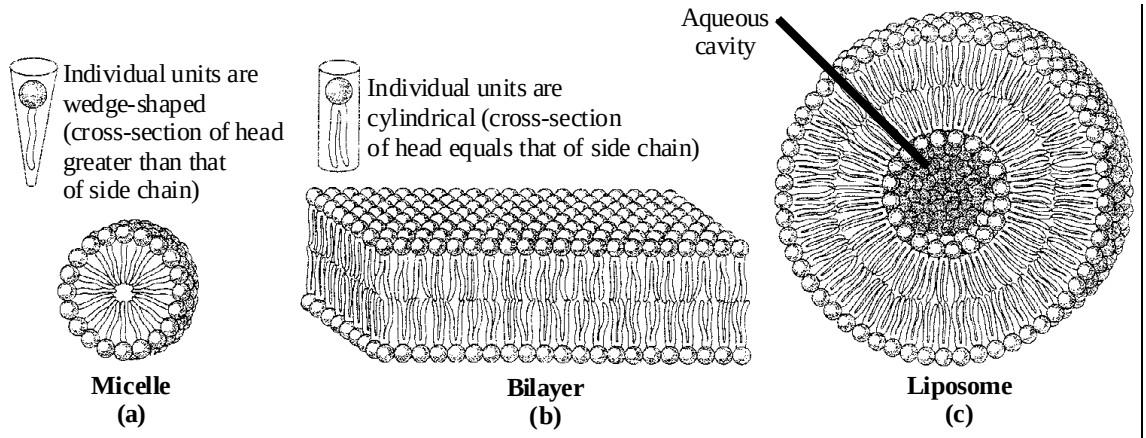
Figure 4: Dipole-dipole interactions. The strength of each dipole is represented by the thickness of the accompanying arrow. (A) Interactions between permanent dipoles. These interactions, here represented by carbonyl groups lined up head to tail, may be attractive, as shown here, or repulsive, depending on the relative orientations of the dipoles. (B) Dipole-induced dipole interactions. A permanent dipole (here shown as a carbonyl group) induces a dipole in a nearby group (here represented by a methyl group) by electrostatically distorting its electron distribution. This always results in an attractive interaction. (C) London dispersion forces. The instantaneous charge imbalance resulting from the motions of the electrons in a molecule (left) induce a dipole in a nearby group (right); that is, the motions of the electrons in neighboring groups are correlated. This always results in an attractive interaction.

London dispersion forces (item c above) contribute to the interaction of all molecules, polar and nonpolar.

D. Hydrophobic interactions

Hydrocarbons are not soluble in water. If you try to mix a purely nonpolar substance with water (such as motor oil), the two components will separate into immiscible liquid phases.

If you mix an amphipathic molecule with water, the hydrophobic portions of the amphiphile will aggregate to minimize their exposure to the solvent. The size and shape of the aggregates depends on the properties of the amphipathic molecule. Fatty acids and other detergents form ***micelles***, globular structures consisting of less than 100 molecules. Phospholipids form bilayer membranes that self-seal to form vesicles called liposomes. (See fig below).



The origin of the hydrophobic effect has both a simple and a complicated explanation.

The simple explanation:

Because water molecules interact so strongly with each other, each water molecule would rather be surrounded by other water molecules than to be partly in contact with a hydrocarbon. To minimize the number of water molecules in contact with hydrocarbon, the solvent actively "squeezes out" the hydrocarbon chains by forcing them to aggregate. (Aggregation reduces their exposed surface area.)

We could express this process in thermodynamic terms as follows. The free energy of water molecules at the hydrocarbon interface is higher than the free energy of water in the bulk phase. Therefore, the aggregation of hydrocarbon chains is favorable because it is accompanied by a net increase in the number of water molecules with lower free energy. (This explanation will have more meaning after the lecture on thermodynamics.)

The complicated explanation (for those who are interested):

The complicated explanation is essentially a more detailed description of the thermodynamics argument presented above. It has been found that when water molecules are forced to be at the hydrocarbon interface, they maximize their interactions with each other to form "ice-like" structures around the "guest" molecule. These ordered structures are energetically favorable, but entropically unfavorable. It is the unfavorable entropy term that makes the free energy of water molecules at the hydrocarbon interface higher that the free energy of water molecules in the bulk phase.

Consequently, when the hydrocarbon chains interact, the entropy of the liberated water molecules increases, leading to a decrease in free energy. For this reason, it is said the hydrophobic interactions are entropically driven.

III. How strong are these forces?

When we look at a protein in its final folded state we often ask: What is the relative importance of each of the forces listed above to maintaining the final structure? Similarly, when we see a molecule bound to protein--a hormone to its receptor, for example--we want to know which of the forces listed above is most important.

We will consider these questions in more detail after learning about binding equilibria and free energy.

Protein Structure

PROTEIN STRUCTURE I & II

INSTRUCTOR: **Saibal Dey, Ph.D.**
Assistant Professor
Department of Biochemistry
Room B4012
Phone: 301-295-3449

READING:
These notes are designed to cover all the essential materials for the two lectures on Protein Structure. **No additional reading is required or suggested**. However, for your information names of a few books that cover this aspect well are listed below.

1) **Lehninger Principles of Biochemistry (Third Edition):** by David Nelson and Michael M. Cox, 2000, Worth Publishers
2) **Biochemistry (Fourth Edition):** by Lubert Stryer, 1995, W.H. Freeman and Company
3) **Introduction to Protein Structure (Second Edition):** by Carl Branden and John Tooze, 1999, Garland Publishing
4) **Proteins, Structures and Molecular Properties (Second Edition):** by Thomas E. Creighton, 1993, W.H. Freeman and Company

LEARNING OBJECTIVES:
After successfully completing these lecture notes you should be able to **address** the following aspects of protein structure

1) Draw a ***polypeptide*** and label the ***main chain*** carbon atoms, the ***N-terminal*** and ***C-terminal*** ends of the polypeptide, the ***side chains*** of the amino acid residues.
2) Draw a ***peptide bond*** and state its conformational flexibility, label ***phi (***ϕ) and ***psi (***ψ***)*** angles.
3) Define the ***structural hierarchy*** of proteins. Describe ***covalent*** and ***non-covalent*** forces that stabilize secondary, tertiary and quaternary structures of proteins. State the ***amino acid residues*** that play important roles in each type of force.

PROTEIN STRUCTURE I & II

1) Describe the structural features of the two major periodic structures of proteins: α ***helix*** and β ***sheet***. Differentiate between them, with special emphasis on the nature of hydrogen bonding.
5) State significance of ***loop regions*** and ***turns*** in protein structure and function.
6) Define ***dipole moment*** of an α helix and state its significance.
7) State the consequence of ***proline residues*** in an α helix, list ***good and poor helix formers*** and state the use of helical wheel projection.
8) Differentiate between ***parallel*** and ***antiparallel*** β ***sheets***.
9) Define ***supersecondary structure*** or ***motifs***, give examples, and mention any specific function associated with each of them.
10) Define ***domains*** and their three main ***types***. List examples of each domain types and subtypes (such as ***Globin fold***, ***coiled coil*** α ***helices***, ***Rossman fold***, ***TIM barrel*** and ***horse shoe fold***) and their functional contributions.
11) List the types of ***posttranslational*** modification of proteins and mention their structural and functional roles, with special emphasis to ***collagen structure***.
12) List important ***metal atoms*** that expand structural and functional ability of proteins, with special emphasis on their role in ***hemoglobin/myoglobin*** and ***zinc fingers***. Name the amino acids whose side chains are excellent metal ligands.
13) Define the basic structural plan of ***Leucine Zipper*** and its functional implication.
14) Describe the principle of ***protein folding*** with reference to ***ribonuclease***.
15) List ***physical*** and ***chemical factors*** that act as protein denaturants and their mode of action.
16) Draw types of ***membrane proteins*** and describe the method of predicting transmembrane regions in a protein from its primary sequence (***Hydropathy plot***).
17) Define protein conformational ***change*** and state how it differs from protein denaturation.

PROTEIN STRUCTURE I & II

WHY WE NEED PROTEINS?

To survive and perpetuate, a living cell must translate and put into action the **genetic information** that is stored in the form of Deoxyribonucleic acid or DNA (in some cases Ribonucleic acid or RNA). This essentially **one-dimensional blueprint** of life is transformed to **three-dimensional molecular activities** inside a living cell by the action of protein molecules. The complex nature of cellular activities demands a vast structural diversity among proteins. The naturally occurring proteins of the present day have evolved to perform specific biological functions, such as

Enzymatic catalysis
Transport and storage
Coordinated motion
Mechanical support
Immune protection
Generation and transmission of nerve impulses
Control of growth and differentiation.

WHY STUDY PROTEIN STRUCTURE?

From biochemical point of view:

One cannot really understand biological reactions without knowing the structural properties of the participating molecules, therefore, the **three-dimensional structures** of the proteins have to be elucidated to assess their true functional properties.

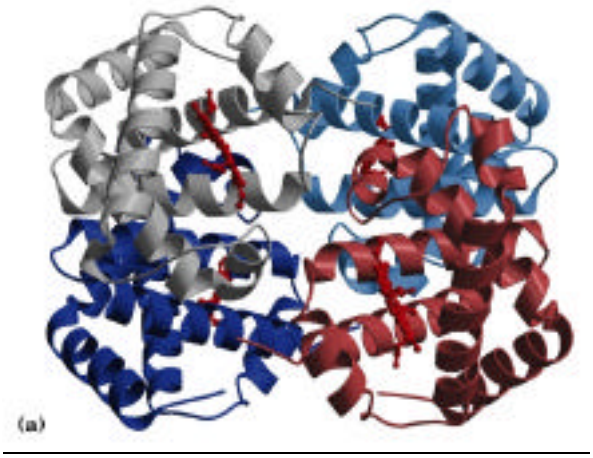


Figure 1. Three Dimensional Structure of Hemoglobin molecule

Myoglobin and **hemoglobin** are the first proteins for which three-dimensional structures were determined, and our current understanding of their function is phenomenal. They, together, illustrate almost every aspect of the most central of biochemical processes: **the reversible binding of a ligand to a protein**.

From Biomedical point of view:

Most disease processes manifest themselves at the level of protein activity. The knowledge on three-dimensional structure of these proteins is invaluable in developing

therapeutic drugs as well as in determining treatment strategy. The discovery of HIV protease inhibitors represents one of the best examples of the application of protein structural knowledge to **rational drug design**.

HIV protease is an enzyme that cleaves polyproteins of the virus into essential functional protein products during the maturation process of the virion. If the polyproteins are not cleaved, the virus fails to mature and is incapable of infecting a new cell. Soon after the discovery of the HIV virus itself, the genetic sequence of the HIV protease enzyme, was determined. Several years later the HIV protease gene product was expressed and crystallized. Once the 3-D structure had finally been worked out, researchers were able to design compounds that specifically targeted the active site of the protease and render it inactive. Like a key perfectly fitted to a lock, the protease inhibitors simply lock up the enzyme and prevent it from functioning. All three clinically approved HIV protease inhibitors, Saquinavir, Ritonavir and Indinavir, are highly effective in keeping the viral load of HIV patients on check and, over the years, had helped cut drastically the number of AIDS related death in developed countries.

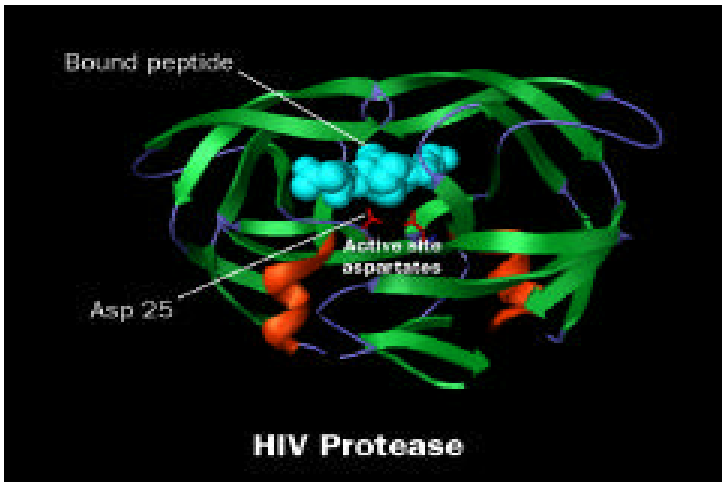


Figure 2 (a). Three dimensional structure of HIV protease

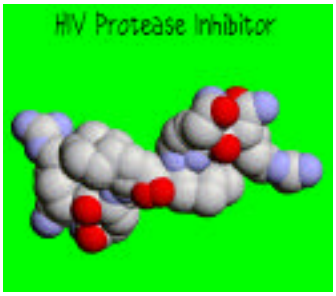


Figure 2 (b). HIV Protease Inhibitor

PROTEIN STRUCTURE I & II

DEFINING MOLECULAR STRUCTURE OF PROTEINS

As we will study in details, proteins are polymers of amino acids held together by covalent attachments in a sequence dictated by the genetic code. The primary sequence of a protein can be determined by either sequencing the gene encoding the protein or by direct chemical analysis of the protein molecule. However, exact prediction of **three-dimensional structure** of proteins from their amino acid sequence is near impossible, instead be determined experimentally. Techniques, like **X-ray crystallography** or **nuclear magnetic resonance** (NMR), are instrumental in elucidating protein structure. Over the past 30 years the structure of more than 6000 proteins have been solved, and the sequences of more than 500,000 have been determined. This has generated a body of information from which a set of **basic principle** of protein structure has emerged. These basic principles help us identify **common structural themes**, relate **structure to function** and see **fundamental relationship** between different proteins. A new era has began in the field of biomedical research where predicting protein structure from its primary sequence and assigning biochemical function from their 3-D structure is becoming increasingly feasible.

CURRENT TREND AND FUTURE DIRECTION

Proteomics:

Since proteins direct virtually every biological function, it makes sense to systematically catalogue all protein molecules expressed by a given **genome** (all the genes in a given organism), and to learn how they change during disease. All the proteins expressed by a given genome are collectively referred to as **proteome** and the science of studying them in details is known as **proteomics**.

Genomics:

Genomics is the study of DNA and the processes, which lead to the creation of proteins. **Functional genomics** is a science that deals with characterizing the biochemical function of all the proteins in a given genome. Since structure predicts function, one promising way to deduce function is by examining three-dimensional structure of gene products; an approach often referred to as **structural genomics**.

PROTEIN STRUCTURE I & II

Bioinformatics:

Bioinformatics is defined as the systematic development and application of computing systems and computational solution techniques analyzing data obtained by experiments, modeling, database search, and instrumentation regarding Biological aspect. Therefore, bioinformatics is a key tool for handling the overwhelming amount

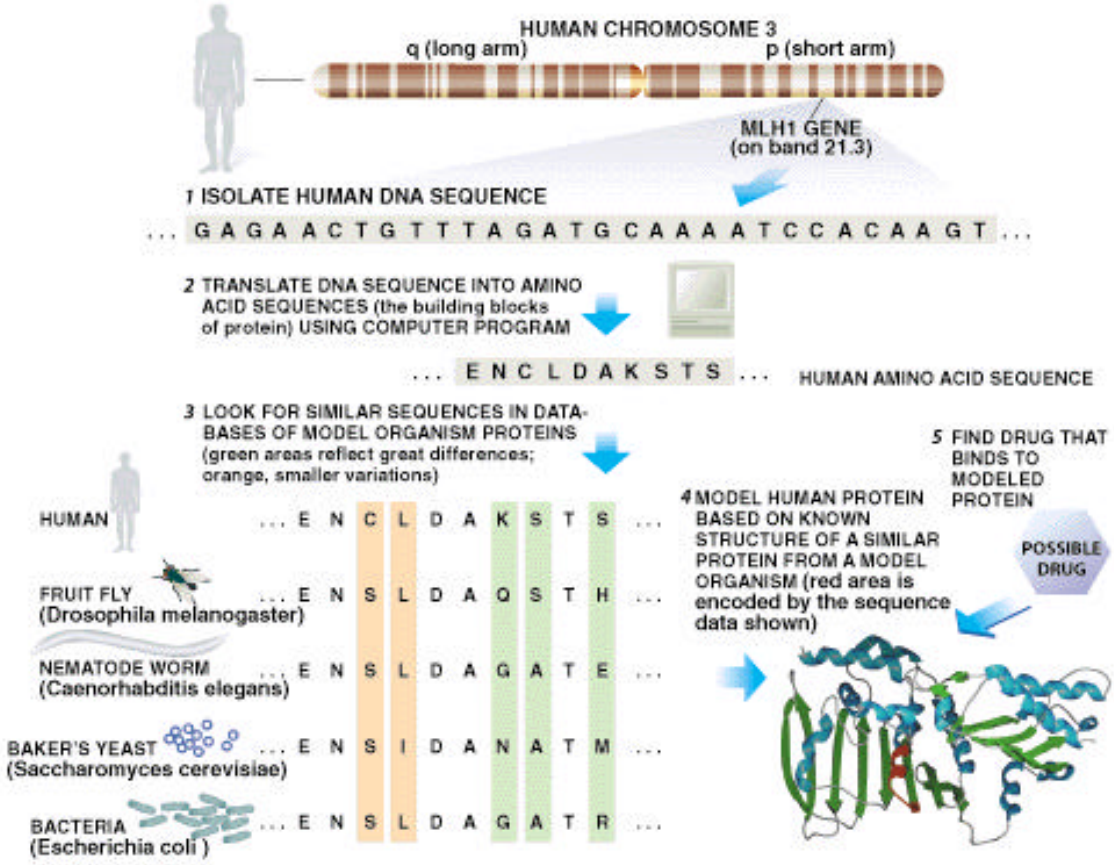


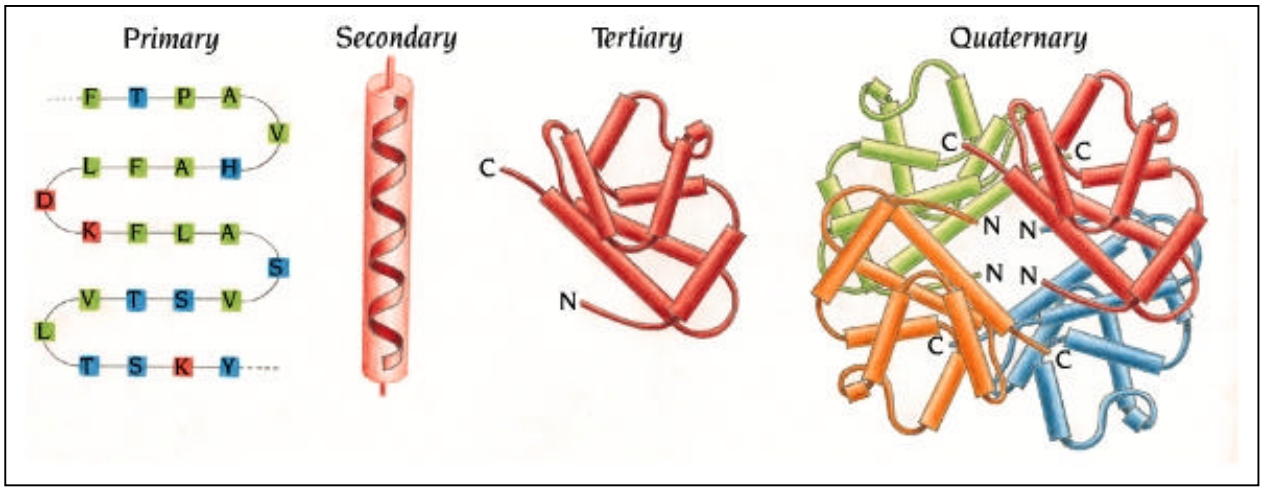
Figure 3. Bioinformatics in understanding protein structure *(Scientific American, July 2000, page 61.)*

of data being generated in the field of structural and functional genomics.

PROTEIN STRUCTURE I & II

PROTEIN MOLECULES ARE ORGANIZED IN A STRUCTURAL HIERARCHY

The three-dimensional structure of protein is complex. The molecular architecture of protein is best understood studying the molecule in terms of four organizational levels, namely**, primary, secondary, tertiary and quaternary.**



STRUCTURAL ORGANIZATION OF PROTEINS

I) PRIMARY STRUCTURE

PROTEINS ARE POLYPEPTIDE CHAINS

1) Amino acids are joined end-to-end during protein synthesis by the formation of **peptide bonds.** The carboxyl group of one amino acid condenses with the amino group of the next eliminating a **water** (H_2O) molecule.
2) The equilibrium of peptide formation lies on the side of hydrolysis rather than synthesis. Hence, biosynthesis of peptide bonds requires an **input of free energy**, whereas their hydrolysis is thermodynamically downhill.
3) The amino group of the first amino acid of a polypeptide chain and the carboxyl group of the last amino acid remain intact and are designated as the **amino** terminus and the **carboxy** terminus ends of the protein, respectively.
4) The formation of succession of peptide bonds generates a "**main chain**" or a "backbone" from which project the various **side chains**.
5) The main chain atoms are $\mathbf{C_\alpha}$ to which the side chain is attached, also are bound to the **NH** group and **C'=O** group of respective amino acid residue. These units or **residues** are linked into a **polypeptide** by peptide bond between the **C'** atom of one residue and the **nitrogen atom** of the next.

PROTEIN STRUCTURE I & II

6) Most polypeptides contain between **50 and 2000 amino acid** residues. The **mean**

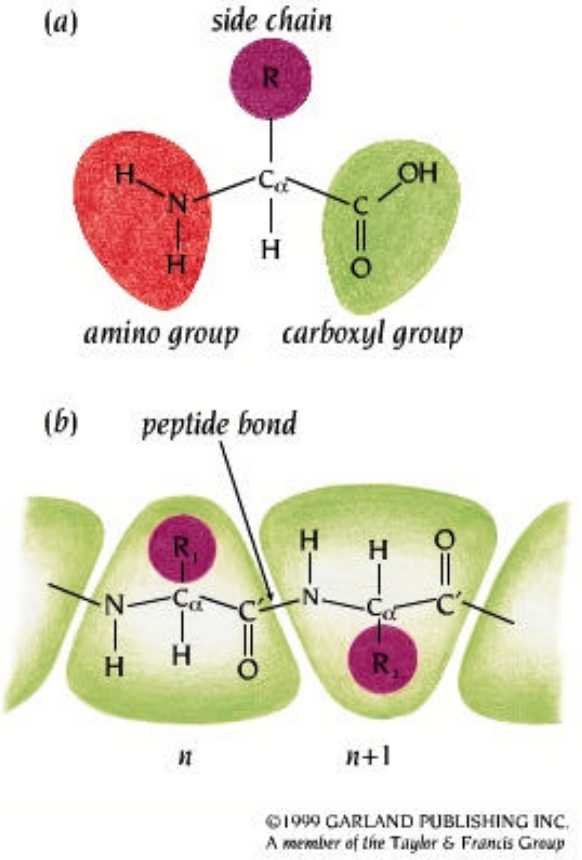
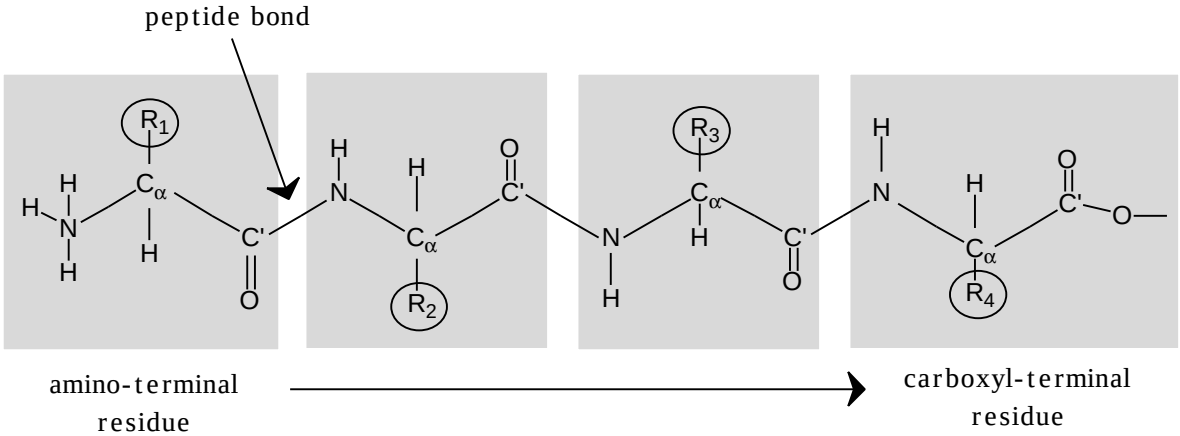


Figure 1.2 (Branden and Tooze). **(a)** Schematic diagram of an amino acid, illustrating the nomenclature used to define the molecule. A central carbon atom (C_α) is attached to an amino group (NH_2), a carboxyl (COOH) group, a hydrogen atom (H) and a side chain (R). **(b)** Schematic representation of a peptide bond between two amino acid residues. In a polypeptide chain the carboxyl group of amino acid "n" has formed a peptide bond, C'-N, to the amino acid "n+1". One water molecule is eliminated in this process. The repeating units, which are called residues, are divided into main-chain atoms and side chains. The main chain part which is identical in all residues, contains a central C_α atom attached to an NH group and C'=O group, and an H atom. The side chain R, which is different for different residues, is bound to the C_α atom.



A tetrapeptide. The constituent amino acid residues are shaded. The chain starts at the amino end.

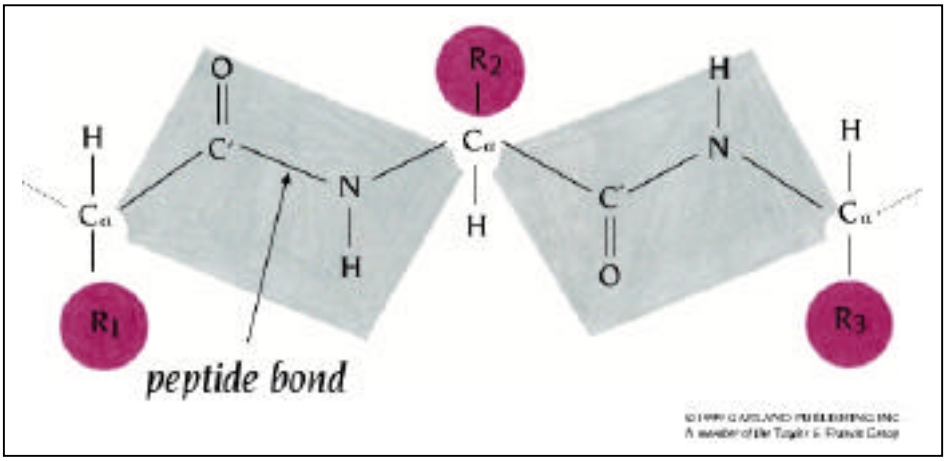
molecular weight of an amino acid residue is about **110**, and so the molecular weights of polypeptide chains are between **5500 and 220,000**. This value is often referred to as the mass of a protein, which is expressed in units of **daltons**; one dalton is equal to one atomic mass unit. A protein with a molecular weight of 50,000 has a mass of 50,000 daltons, or 50 kDa (**kilodaltons**).

7) Proteins have unique amino acid sequences that are genetically determined. This is often referred to as the **primary sequence** or **primary structure** of the protein.

PEPTIDE UNITS ARE BUILDING BLOCKS OF PRTOEIN STRUCTURES

8) Peptide units are effectively **rigid groups** that are linked into a chain by covalent bonds at the C_α atoms; the only **degrees of freedom** they have are rotations around the **C_α-C'** and the **N-C_α** bonds**.**

9) By convention the angle of rotation around the N-C_α bond is called **phi (ϕ)** and the angle around the C_α-C' bond from the same C_α atom is called **psi (ψ)**.



Part of a polypeptide showing peptide units. Each peptide unit contains the C_α atom and the C'=O group of the residue n as well as the NH group and the C_α atom of residue n+1. Each such unit is a planar, rigid group with known bond distances and angles. R_1, R_2 and R_3 are the side. The peptide group is planar because rotation around the C'-N bond is prevented.

10) Most conformations of ϕ and ψ angles for an amino acid are not allowed because of steric **collisions** between the **side chains** and **main chain**.

11) The **rigidity** of the peptide bond enables protein to have **well-defined** three-dimensional forms. On the other hand, the **freedom of rotation** on the either side of the peptide unit is equally important because it allows proteins to **fold in many different ways**.

12) The folded conformations of proteins need to be **stabilized** for functional activity.

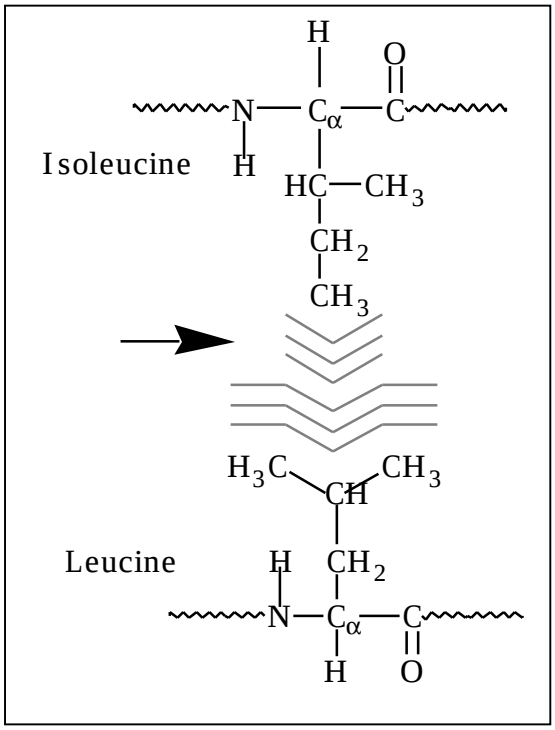
PROTEIN STRUCTURE I & II

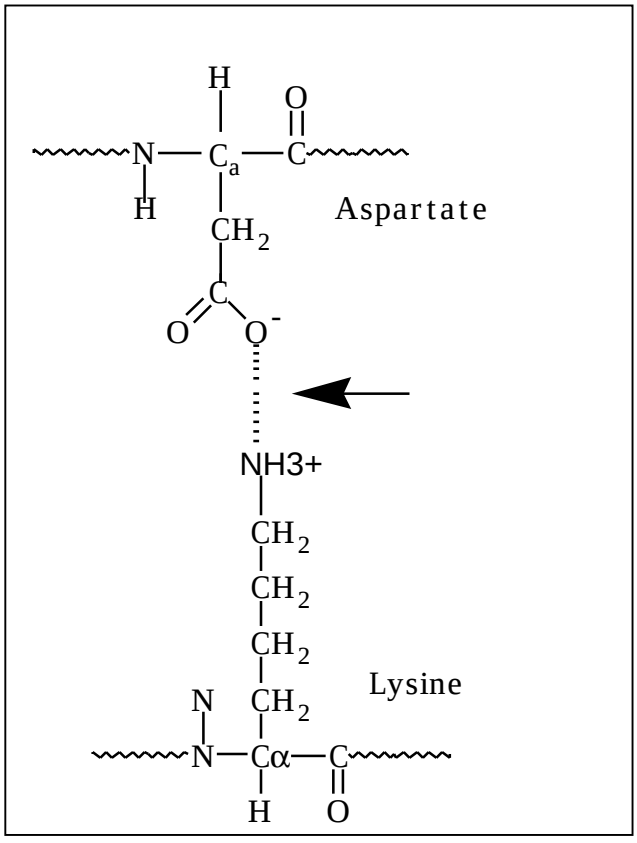
FORCES STABILIZING THREE-DIMENSIONAL STRUCTURE

Molecular interactions among the amino acid residues within a polypeptide chain and between the chain and its immediate environment drive the three-dimensional structure of a protein molecule. All reversible molecular interactions in biological systems are mediated by four kinds of forces, **hydrophobic interaction**, **electrostatic bonds**, **hydrogen bonds,** and **van der Waals interaction**.

Hydrophobic interaction:

In a polar environment, amino acids with non-polar side chains tend to be located in the interior of the polypeptide molecule, where they associate with other hydrophobic amino acid residues. In contrast, amino acids with polar or charged side chains tend to be located on the surface of the molecule in contact with the polar solvent. Proteins located in nonpolar environments, such as, membrane proteins exhibit the reverse arrangement



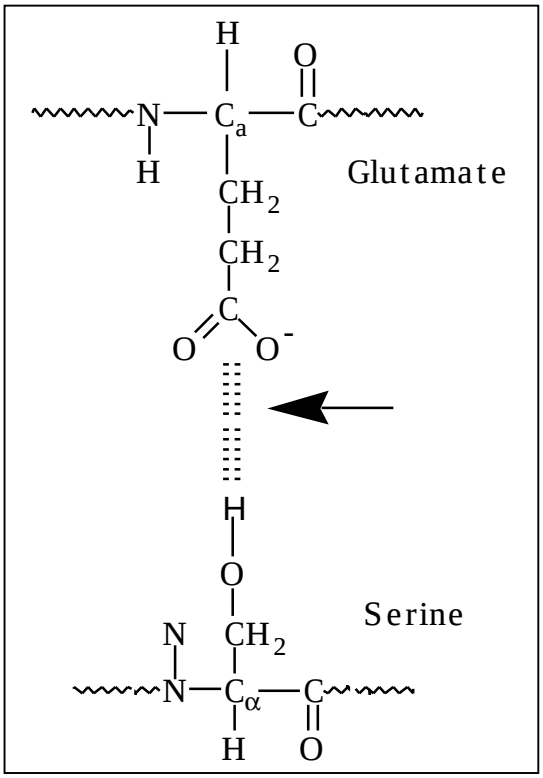


Electrostatic bonds:

Negatively charged groups, such as carboxyl group ($-COO^-$) in the side chain of aspartate or glutamate can interact with positively charged groups, such as the amine ($-NH_3^+$) in the side chain of lysine when present in close proximity to each other.

PROTEIN STRUCTURE I & II

Hydrogen bonds:



Proteins are rich in hydrogen bonding potentiality. **Side chains** of 11 out of 20 fundamental amino acids, together with the **main chain** NH and C=O groups can participate in hydrogen bonding.

1) These amino acids can be grouped according to their hydrogen bonding potentialities. The side chain of tryptophan (W) and arginine (R) can serve as hydrogen bond **donors only**.
2) The peptide unit itself and the side chain of asparagine (N), glutamine (Q), serine (S) and threonine (T) can serve as hydrogen bond **donors or acceptors**.
3) The hydrogen bonding capability of lysine (K), the terminal amino group, aspartic acid (D), glutamic acid (E), the terminal carboxyl group, tyrosine (Y) and histidine (H) vary with pH. These groups can serve as **acceptors and donors** within a certain pH range, and as **acceptors or donors** (but not both) at other pH values. The hydrogen-bonding modes of these ionizable residues are **pH-dependent.**

II SECONDARY STRUCTURE

INTERIOR OF THE PROTEIN IS HYDROPHOBIC

5) Water-soluble globular protein molecules pack hydrophobic side chains into the interior of the molecule, creating a **hydrophobic core** and a **hydrophilic surface**.
6) The hydrophilic nature of the main chain which is highly polar with one **hydrogen bond donor**, **NH**, and one **hydrogen bond acceptor**, **C'=O**, for each peptide unit, creates problem in forming such a hydrophobic core.
7) In a hydrophobic environment, these main chain polar groups must be **neutralized** by the formation of hydrogen bond. This results in formation of **regular secondary structure** within the interior of the protein molecule.
8) Such secondary structures are usually of two types:
   - **A) Alpha helices** (α helices)
   - **B) Beta sheets** (β sheets)

PROTEIN STRUCTURE I & II

Both types are characterized by repetitive **hydrogen bonding** between the main-chain NH and C'=O groups.

A non-repetitive structure constitute the third type:

C) Loops (loop regions)

A) The alpha (α) helix

a) Alpha helix is a rod-like structure, in which the tightly coiled polypeptide main chain forms the inner backbone, and the side chains are extended outward in a **helical array**.

b) It is formed by a stretch of consecutive amino acid residues that have ϕ **and** Ψ **angle** pair approximately **–60° and –50°**.

c) α helix is stabilized by hydrogen bond formation between **C'=O of residue n** and **NH of residue n + 4**. Thus all NH and C'=O groups are joined with hydrogen bond except the first NH group and the last C'=O groups at the ends of the α helix.

d) α helices vary in size from four to over 40 residues. The average length is around **10 residues** per helix, which corresponds to **three helical turns**.

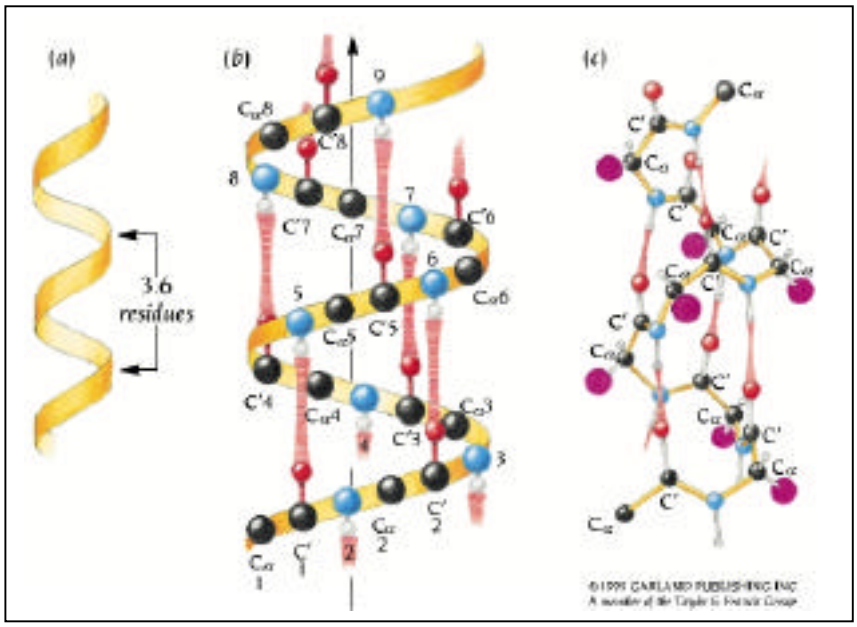


Figure 2.2. (Branden and Tooze) Models of right-handed α helix. Main-chain N and O atoms are hydrogen-bonded to each other within α helices. (a) Idealized diagram of the path of the main chain in an α helix. Alpha helices are frequently illustrated in this way. (b) The same as (a) but with approximate positions for main-chain atoms and hydrogen bonds included. The arrow denotes the direction from the N-terminus to the C-terminus. (c) Schematic diagram of an α helix.

e) Each residue is related to the next one by a rise of **1.5Å** along the helical axis, this is also called as the translation per residue.

f) The rotation per residue is **100°**, which gives **3.6 residues** per turn of helix.

g) The **pitch** of the α-helix, which is equal to the product of

the translation (1.5 Å) and the number of residues per turn (3.6), is **5.4 Å.**

h) In theory, an α helix can be either **right-handed** or **left-handed** depending on the screw direction of the chain. The type of α helix that is observed in proteins is almost always **right-handed**.

The α helix has a dipole moment

i) Since the peptide units are aligned in the same orientation along the helical axis, all hydrogen bonds in an α helix point in the same direction.

j) Each peptide unit has a **dipole moment** arising from the different polarity of NH and C'=O groups and these dipole moments are also aligned along the α helical axis.

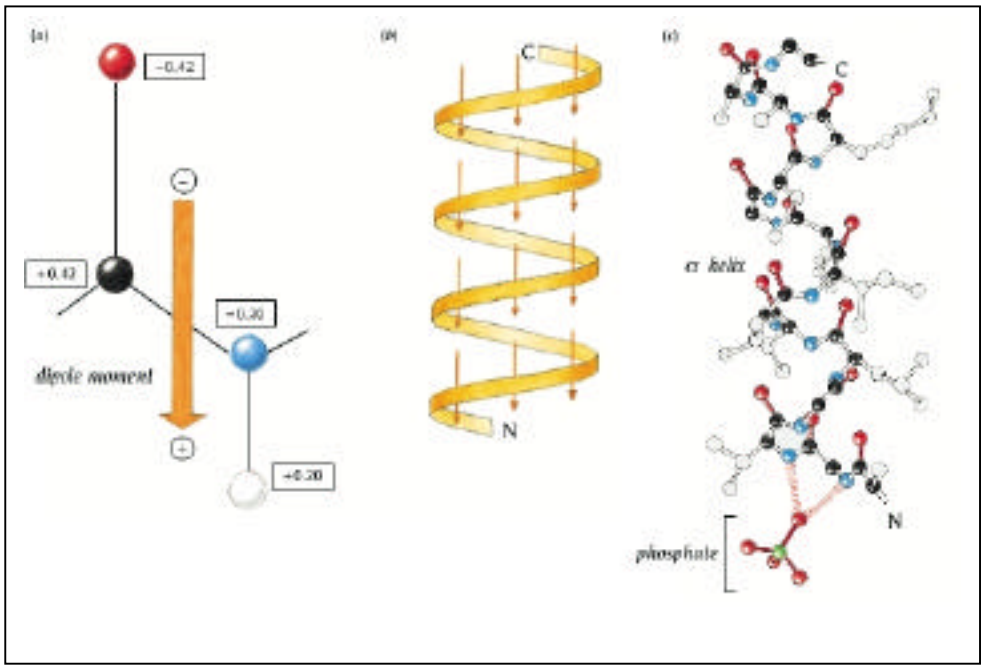


Figure 2.3 (Branden & Tooze). (a) The dipole of a peptide unit. Values in boxes give approximate fractional charges of the atoms of a peptide unit. **(b)** The dipoles of peptide units are aligned along the α helical axis, which creates an overall dipole moment of the α helix, positive at the amino-terminal end and negative at carboxy end. **(c)** a phosphate group hydrogen-bonded to the NH end of the α helix.

k) Consequently, the overall effect is a significant **net dipole** for the α helix that gives a partial **positive charge** at the **amino end** and a partial **negative charge** at the **carboxy end** of the α helix.

l) The dipole moment of an α helix as well as hydrogen bonding potential of the free NH groups at the end of the helix favors binding of negatively charged groups such as **phosphate ions**.

Some amino acids are preferred in α helices

m) The amino acid side chains project out from the α helix and do not interfere with its formation, except for **proline**. The last atom of the proline side chain is bounded to the main-chain N atom, which forms a **ring structure**, C_α-CH_2- CH_2-

CH_2-N. This prevents the N atom from participating in hydrogen bonding and also provides some **steric hindrance** to the α helical conformation. Proline fits very well in the first turn of an α helix, but usually produces a **significant kink** if it is anywhere else in the helix.

n) Different **side chains** have been found to have weak but **definite preferences** either for or against being in α helices. Thus Ala (A), Glu (E), Leu (L), and Met (M) are **good** α **helix formers**, while Pro (P), Gly (G), Tyr (Y), and Ser(S) are **very poor**.

o) Alpha helices can be either completely **buried** within the protein or completely **exposed**.

p) When α helices are along the outside of the protein with one side of the helix facing the solution and the other side facing the hydrophobic interior there is a tendency for side chains to change from **hydrophobic to hydrophilic** with a periodicity of 3-4 residues.

Helical wheel projection

q) A convenient way to illustrate the amino acid sequences in helices is the **helical wheel** or spiral.

r) Since one turn in α helix is 3.6 residues long, each residue can be plotted every **360/3.6 = 100°** around a circle or a spiral.

s) Such a plot shows the **projection of the position** of the residues onto a plane perpendicular to the helical axis.

t) This kind of analysis helps identifying putative α helices with **amphipathic** property (one side of the helix is hydrophilic and the other is hydrophobic).

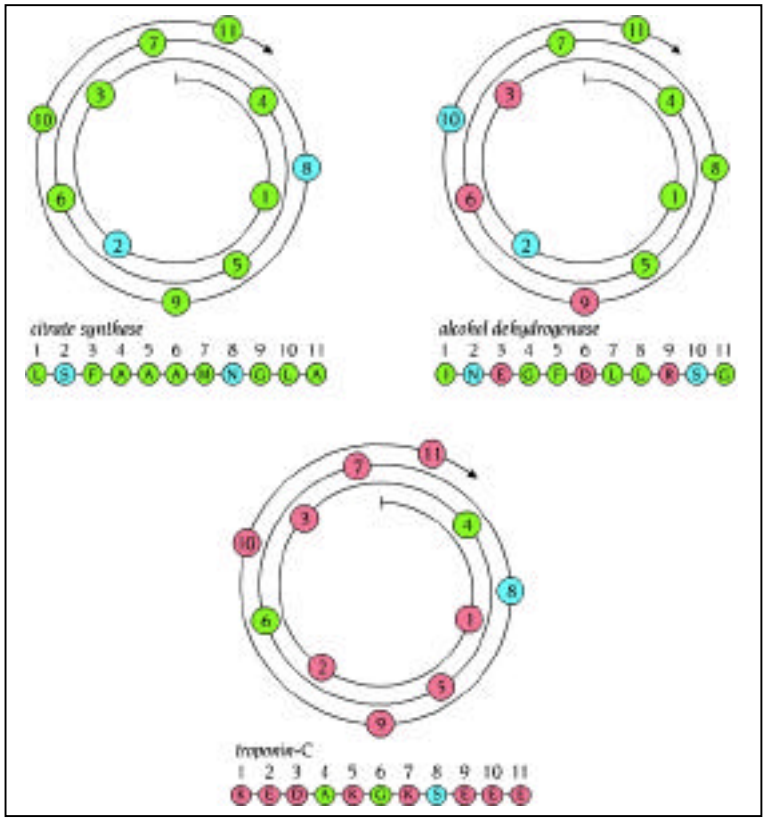


Figure 2.4 (Branden & Tooze). The helical wheel or spiral. Amino acid residues are plotted every 100° around the spiral, following the sequences shown below.

u) Alpha helices **spanning membranes** are in a **hydrophobic** environment. Therefore, most of their side chains are hydrophobic.

B) The Beta (β) sheets

a) The second major structural element found in globular protein is the β **sheet**.

b) This structure is built from a combination of **several regions** of the polypeptide chain, in contrast to α helix, which is built up from one continuous region. These regions, β strands, are usually **5-10 residues long.**

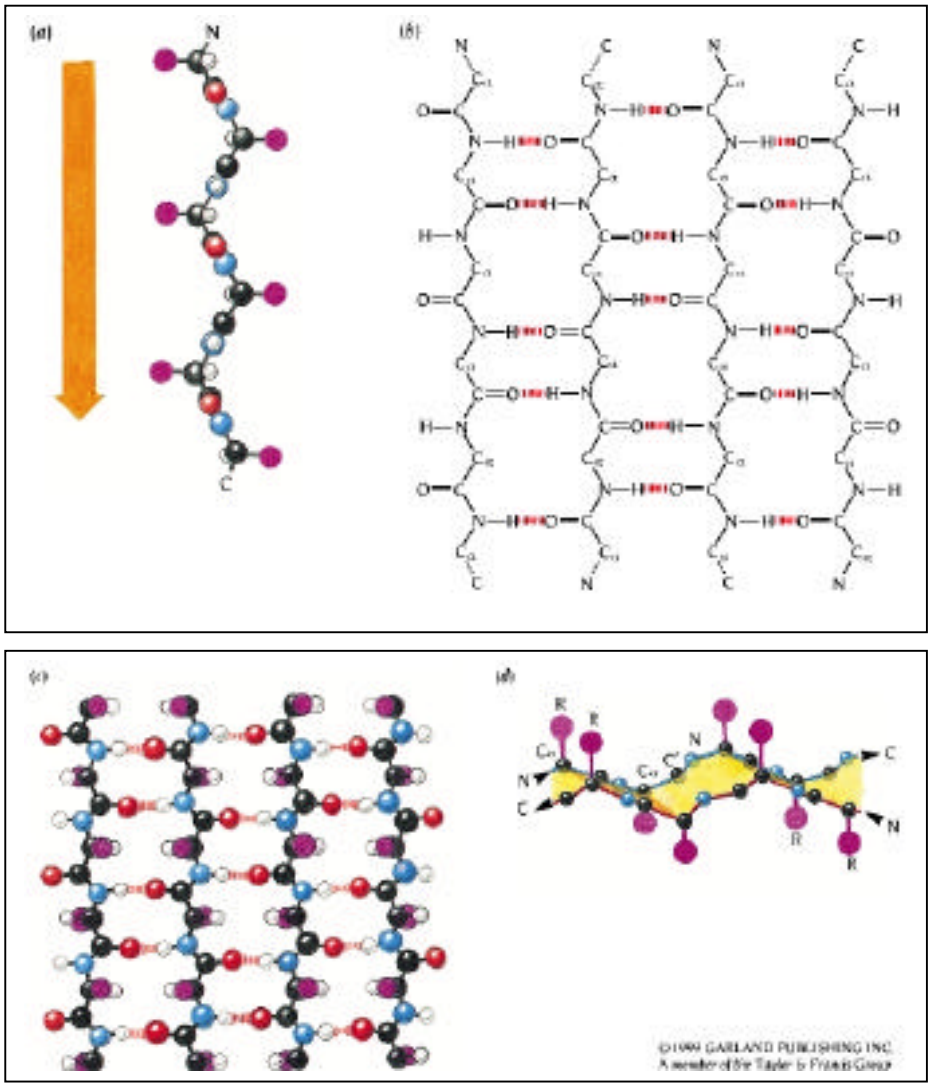


Figure 2.5 (Branden & Tooze). Schematic illustrations of antiparallel β sheets. **(a)** The extended conformation of a β strand. The orientation of the β strand is at right angles to those of (b) and (c). A β strand is schematically illustrated as an arrow, from N to C terminus. **(b)** Schematic illustration of the hydrogen pattern in an antiparallel β sheet. Main-chain NH and O atoms within a β sheet are hydrogen bonded to each other. **(c)** A ball-and-stick version of (b). **(d)** Illustration of the pleat of a β sheet. Not that the directions of the side chains, R, follow the

c) These β strands are **aligned adjacent to each other** such that hydrogen bonds can form between C'=O groups of one β strand and the NH groups on an adjacent β strand and vice versa.

d) The β sheets that are formed from several such strands are "**pleated**" with C_α atoms successively a little above and below the plane of the β sheet. The side chains follow this pattern such that within a β strand they also point alternately above and below the β sheet.

e) Beta strands can interact in two ways to form a pleated sheet, **parallel** and **antiparallel** β sheet.

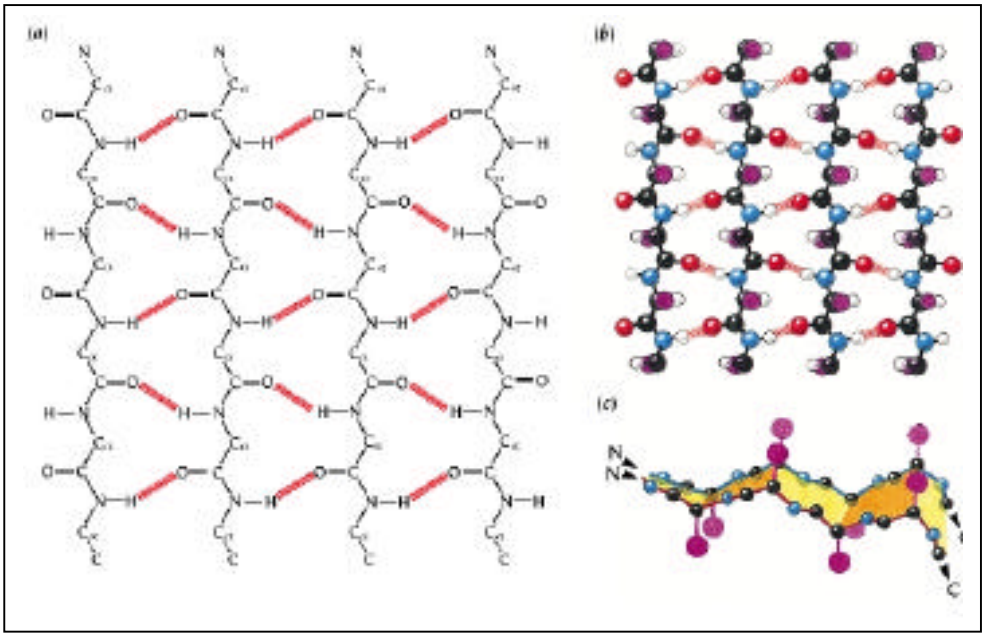


Figure 2.6 (Branden & Tooze). Parallel β sheet. (a) Schematic diagram showing the hydrogen pattern in a parallel β sheet. (b) Ball-and-stick version of (a). (c) Schematic diagram illustrating the pleat of a parallel β sheet.

f) Antiparallel β sheet has **narrowly spaced** hydrogen bond pairs that alternate with **widely spaced** pairs. Parallel β sheets have **evenly spaced** hydrogen bonds that bridge the β strands at an angle. In both types of β sheet all possible main-chain hydrogen bonds are formed, except for the two **flanking strands**.

g) β strand can also combine into **mixed** β **sheets** with some β strand pairs parallel and some antiparallel. There is a strong bias against mixed sheets; only about 20% of the strands inside the β sheets of known protein structures have parallel bonding on one side and antiparallel bonding on the other.

h) As they occur in known protein structures, almost all β sheets, parallel, antiparallel, and mixed, have **twisted strands**. This twist always has the same handedness, which is defined as a **right-handed twist**.

C) Loop regions

a) When present in a protein, α helices and β sheets are connected by **loop regions** of various lengths and shapes.

b) The loop regions are generally at the **surface** of the molecule. Loop regions exposed to solvent are rich in **charged** and **polar hydrophilic residues**.

c) The C'=O and NH groups of the loop regions, when exposed to the solvent, may **hydrogen bond with water molecules**.

d) Loop regions frequently participate in forming **binding sites** (for ligands) and enzyme **active sites**. The antigen binding sites in antibodies are built up from six loop regions.

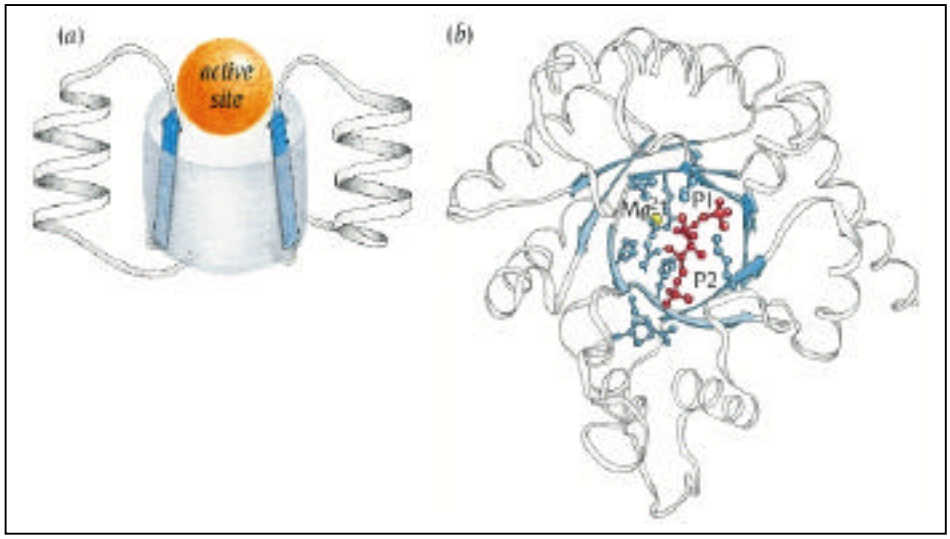


Figure 4.8 (Branden & Tooze). The active site in all α/β barrels is in a pocket formed by the loop regions that connect the carboxy ends of the β strands with the adjacent a helices, as shown schematically in **(a),** where only two such loops are shown. **(b)** A view from the top of the barrel of the active site of the enzyme RuBisCo (ribulose bisphosphate carboxylase), which is involved in CO_2 fixation in plants.

e) These loop regions have **preferred structures** and are not random collection of possible structures.

f) Loop regions that connect two adjacent antiparallel β strands are called **hairpin loops**. Short hairpin loops are called **reverse turns** or simply turns.

g) Long loop regions are often flexible and can

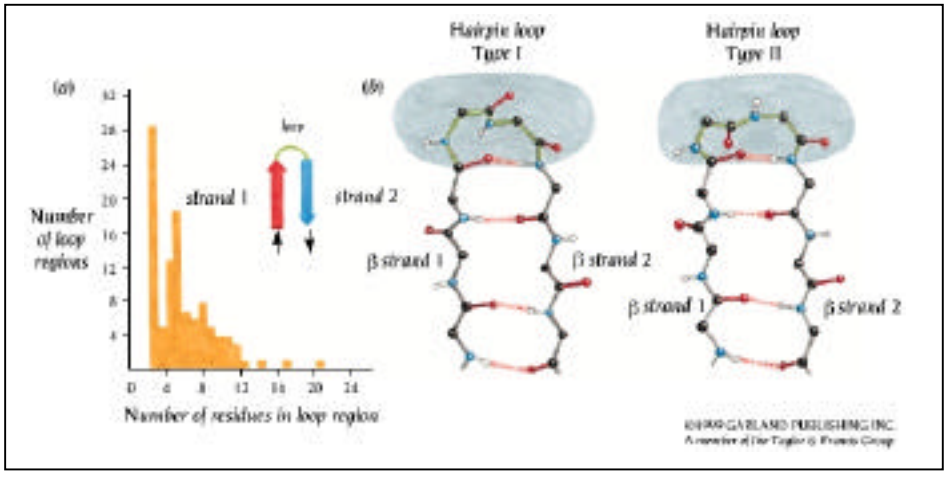


Figure 2.8 (Branden & Tooze). Adjacent antiparallel β strands are joined by hairpin loops. Such loops are frequently short and do not have regular secondary structure. **(a)** Histogram showing the frequency of hairpins of different lengths in 62 different proteins. **(b)** the two most frequently occurred hairpin loops; Type I turn to the left and Type II turn to the right.

frequently adopt several different conformations that are frequently **involved in function** of the protein.

h) During evolution, core structures are much more conserved than loops.

Note: the structure of a protein can be symbolized in highly schematic form by representing β strands as broad **arrows** and, α helices as **helical ribbons**, and connecting regions as **strings**.

MOTIFS OR SUPERSECONDARY STRUCTURES

Simple combinations of a few secondary structure elements with a specific geometric arrangement have been found to occur frequently in protein structures. These units have been called either **supersecondary** structures or **motifs**. These motifs could either be associated with a **particular function** or are part of larger structural or functional assemblies.

Examples:

A) Helix-loop-helix motif

a) Helix-loop-helix is the **simplest motif** with a specific function. It consists of **two α helices joined by a loop region**.

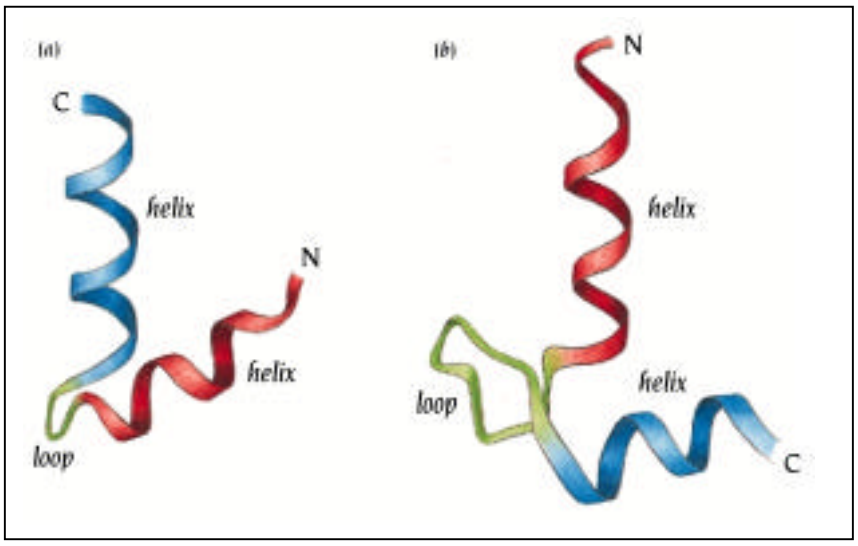


Figure 2.12 (Branden and Tooze). Two α helices that are connected by a short loop region in a specific geometric arrangement constitute a helix-turn-helix motif **(a),** and the calcium-binding motif **(b),** which is present in many proteins whose function is regulated by calcium.

b) One of these motifs, also called the **helix-turn helix motif**, is specific for **DNA binding**. A tandem repeat of two DNA-binding domains, called the **POU** region, is present in a class of eukaryotic transcription factors.

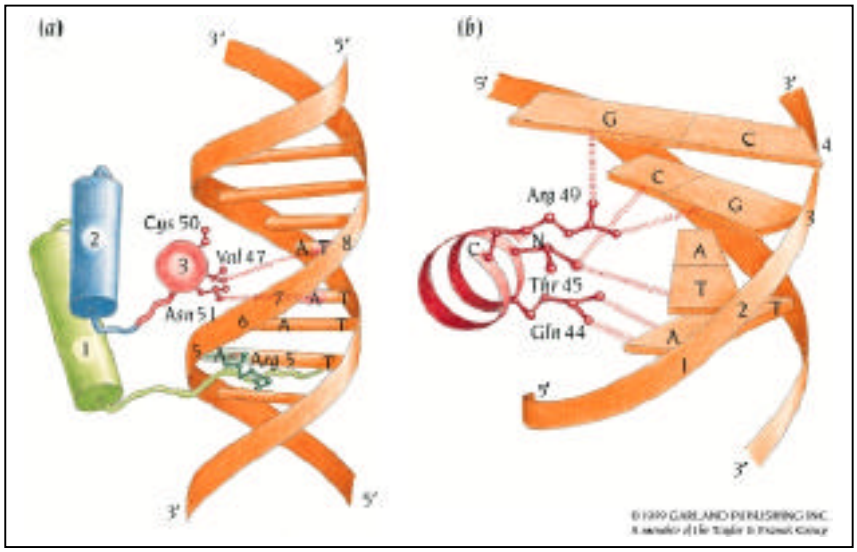


Figure 9.15 (Branden & Tooze). Diagram illustrating the sequence-specific contacts between DNA and the helix-turn-helix motif of POU region. **(a)** Contacts from the POU homeodomain, three residues from the recognition helix interact with base pairs in the major groove of the second half of the POU-binding site. **(b)** Contacts from the POU-specific region. Three residues from the recognition helix, interact with bases in the major groove of the first half of the POU-binding site.

c) Another one is specific for **calcium binding** and is present in paravalbumin, calmodulin, troponin-C, and other proteins that bind calcium and thereby regulate cellular activities. The helix-loop-helix motif provides a **scaffold** that holds the **calcium ligands** in the proper position to bind and release calcium.

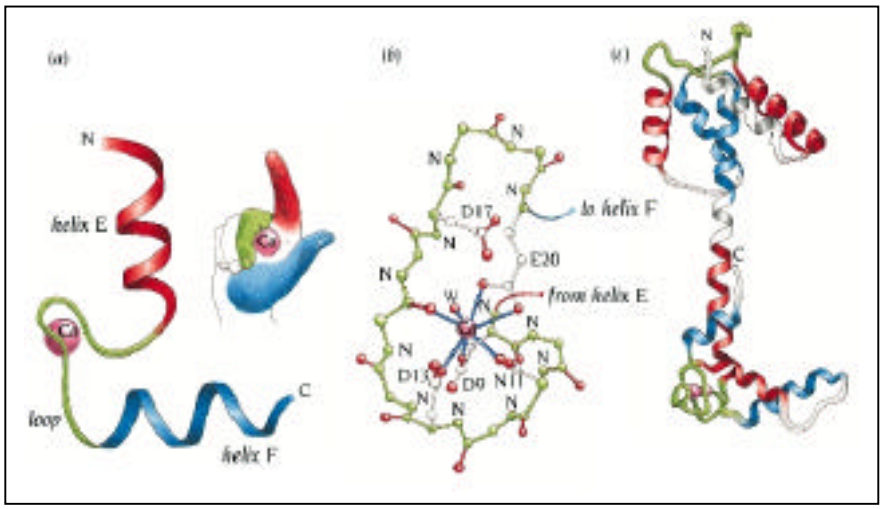


Figure 2.13 (Branden & Tooze). Schematic diagrams of the calcium-binding motif. **(a)** The calcium-binding motif is symbolized by a right hand. Helix E runs from the tip to the base of the forefinger. The flexed middle finger corresponds to a region of 12 residues that bind calcium. Helix F runs to the end of the thumb. **(b)** The calcium atom is bound to one of the motifs in muscle protein toponin-C through six oxygen atoms. In addition, a water molecule (W) is bound to the Calcium ion.

B) Hairpin β motif

a) The **simplest motif involving β strands.** It consists of **two adjacent antiparallel strands** joined by a loop.

b) This motif, called **hairpin** or **β-β unit**, occurs quite frequently in antiparallel β strands.

c) There is **no specific function** associated with this motif.

C) The β-α-β motif

a) When two adjacent β strands are consecutive in the amino acid sequence, the two ends that must be joined are at opposite edges of the β sheet.

Figure 2.17 (Branden & Tooze). Two adjacent parallel β strands are usually connected by an α helix from the C-terminus of strand 1 to the N-terminus of strand 2. Most protein structures that contain β sheets are built up from combinations of such β–α–β motifs. **(a)** Schematic diagram of the path of the main chain. **(b)**Topological diagrams of the β–α–β motif.

b) The polypeptide chain in between the two strands **traverse the β sheet** from one edge to the other and connect the next β strand close to the point where the first β strand started.

c) Such crossover connections are frequently made by α **helices**.

This motif is called a **beta-alpha-beta (β-a-β) motif** and is found as a part of almost every protein structure that has a parallel β sheet.

III) TERTIARY STRUCTURE

LARGE POLYPEPTIDE CHAINS FOLD INTO SEVERAL DOMAINS

18) The fundamental unit of **tertiary structure** is the **domain**.

19) A domain is defined as a polypeptide chain or a part of a polypeptide chain that can **fold independently** into a stable tertiary structure.

20) Domains are also **units of function**. Often the different domains of a protein are associated with different functions.

21) Proteins may be comprised of a **single** domain or as many as **several domains**.

22) There are no fundamental distinction between **domain** and a **subunit**. There are many known examples where several biological functions that are carried out by separate polypeptide chains in one species are performed by domains of a single protein in another species.

23) On the basis of simple considerations of connected motifs, taxonomy of protein structure has classified domain structures into three main groups:

A) Alpha-Domain (α-domains)

B) Beta-domains (β-domains)

C) Alpha/Beta-domains (α/β domains)

PROTEIN STRUCTURE I & II

A) Alpha-Domain structures

In α structures the core is built up exclusively from α helices.

Examples:

i) The globin fold

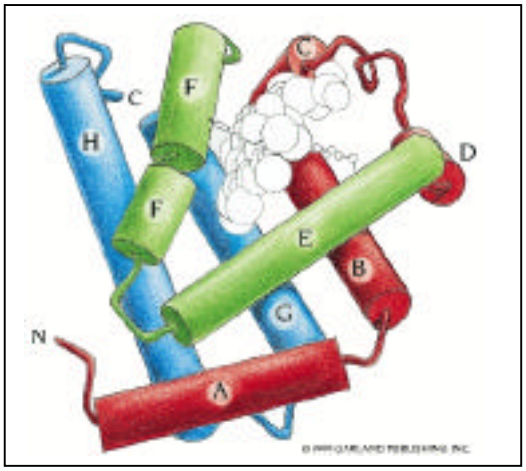


Figure 3.10 (Branden & Tooze). Schematic diagram of the globin domain. The eight α helices are labeled A-H. The heme group is shown in white.

a) One of the most important α structures is the **globin fold**.

b) This fold is present in a large group of related proteins, including **myoglobin**, **hemoglobins** and light capturing assemblies in algae, the **phycocyanins**.

c) Short α **helices** are connected by **loop** regions and packed together to produce a **hydrophobic core**. The α**-helices** constitute the active site, which in myoglobin and hemoglobins bind a **heme group**.

d) Packing interaction within the core hold the helices together in a stable globular structure, while the **hydrophilic residues on the surface** make the protein soluble in water.

ii) Coiled-coil α helices

a) Two α helices together may wound around each other to form a left-handed **supercoil**. This is called the **coiled-coil** arrangement.

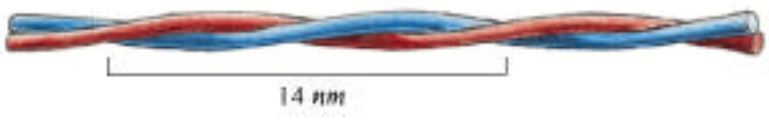


Figure 3.1 (Branden & Tooze). A helical coiled coil. The two α helices wind around each other to form a super helix.

b) Coiled-coils are the basis for some of the fibrous proteins, where the structure may extend over many hundreds of amino acid residues to produce long, flexible dimers that contribute to the strength of the fibers.

c) Left-handed supercoil of two right-handed α helices reduces the number of residues per turn in each helix from 3.6 to 3.5 so that the pattern of side chain interactions between the helices repeats **every seven residues** (after two helical turns).

d) Such sequences are repetitive with a period of seven residues, called the **heptad repeat**. Therefore heptad repeats provide strong indications of α helical coiled-coil structures.

PROTEIN STRUCTURE I & II

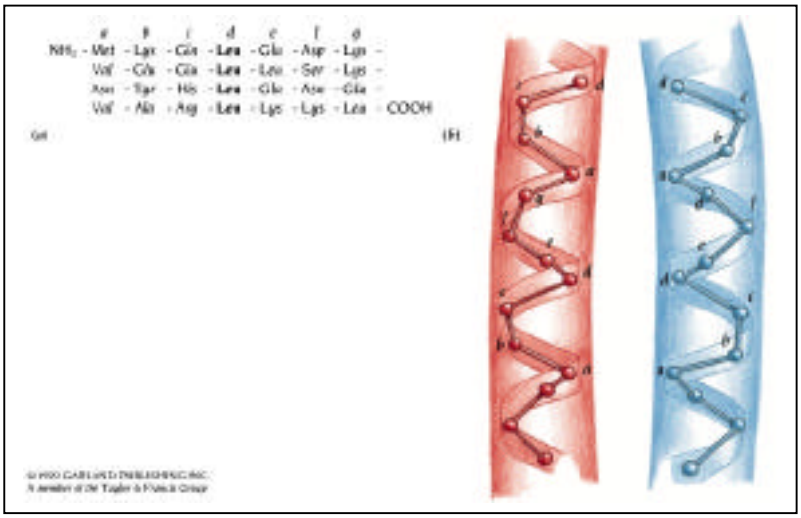


Figure 3.2 (Branden & Tooze). Repetitive pattern of amino acids in coiled-coil α helix. **(a)** The amino acid sequence of the transcription factor GCN4 showing a heptad repeat of leucine residues. Within each heptad the amino acids are labeled a-g. **(b)** Schematic diagram of one heptad repeat in a coiled-coil structure showing the backbone of polypeptide chain.

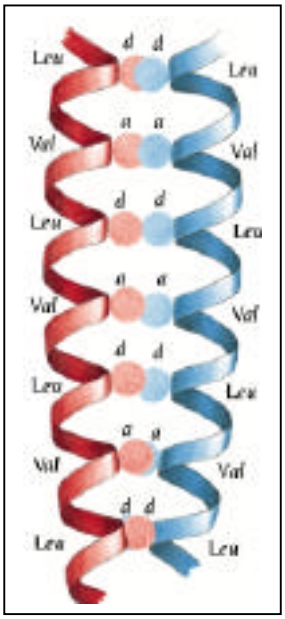


Figure 3.3 (Branden & Tooze). Schematic diagram showing the packing of hydrophobic side chains between the two α helices in a coiled-coil structure. Every seven residues in both α helices is a leucine, labeled "d." Due to the heptad repeat, the d-residues pack against each other along the coiled coil. Residues labeled "a" are also usually hydrophobic and participate in forming the hydrophobic core along the coiled coil.

e) They have been found to be present in different proteins with very diverse functions. For example, **Fibrinogen** (which plays an essential role in blood coagulation), **collectins** (a class of cell surface recognition proteins), some RNA- and DNA-binding proteins, and in both **spectrin** and **dystrophin** (which link actin molecules and the muscle protein myosin) all contain coiled-coil α helices.

B) Beta domain structures

a) Antiparallel β structures comprise the most **functionally diverse** group of protein domain structures. Proteins containing this structure includes enzymes, transport proteins, antibodies and cell surface proteins.

b) The cores of these domains are built up of β **strands** that are arranged in a predominantly antiparallel fashion in such away that they form two β sheets that are joined together.

c) The β sheets are **twisted** and form a **barrel–like** structure.

d) In general they have a hydrophobic core made up of side chains of the β strands. The surface is formed of residues from the loop regions and from the strands.

PROTEIN STRUCTURE I & II

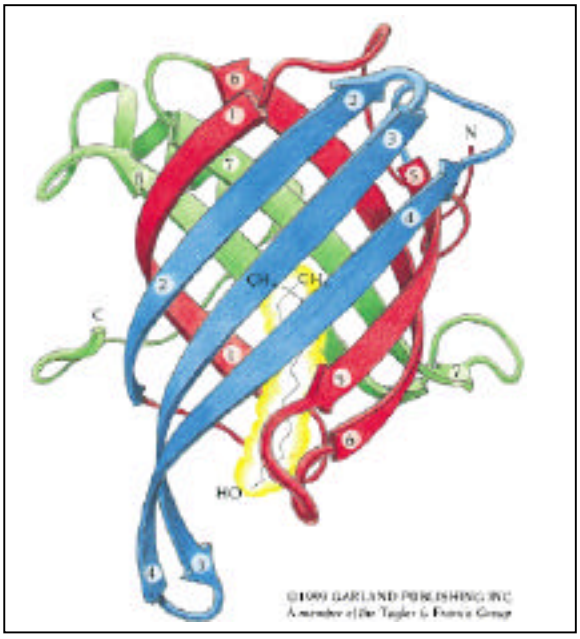


Figure 5.3 (Branden & Tooze). Schematic diagram of the structure of human plasma retinol-binding protein (RBP), which is an up-and-down β barrel. The eight antiparallel β sheets twist and curl such that the structure can also be regarded as two β sheets packed against each other. A retinol molecule, vitamin A, is bound inside the barrel, between the two β sheets, such that its only hydrophilic part (an OH tail) is at the surface of the molecule.

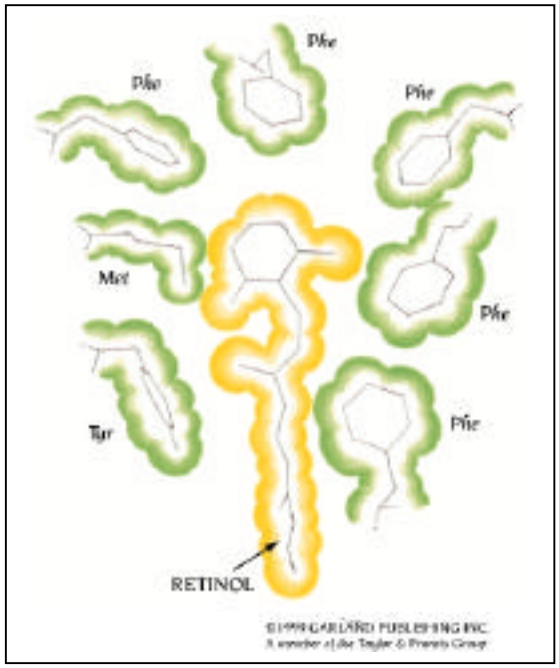


Figure 5.4 (Branden & Tooze). The binding site for retinol inside the RBP barrel is lined with hydrophobic residues. They provide a hydrophobic surrounding for the hydrophobic part of the retinol molecule.

PROTEIN STRUCTURE I & II

C) Alpha/Beta domains (also called Alpha/Beta barrel)

The most frequent of the domain structures are the α/β **domains**. It consists of a central **parallel or mixed** β **sheet** surrounded by α **helices**. In α/β domains, **binding crevices** are formed out of **loop regions**. These regions do not contribute to the structural stability of the fold but participate in **binding and catalytic** action.

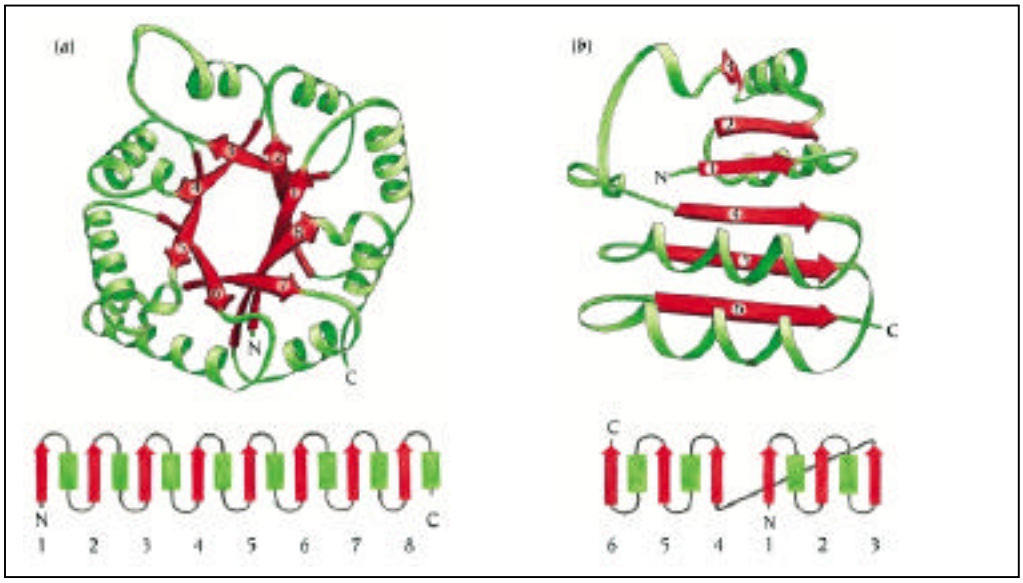
There are **three main classes** of α/β proteins

i) TIM barrel.

a) A core of twisted parallel β strands arranged close together, like the staves of a barrel.

b) The α helices that connect the parallel β strands are on the outside of the barrel.

c) This domain structure is often called the **TIM barrel** from the structure of the enzyme **T**riosephosphate **I**so**m**erase, where it was first observed.



ii) Rossman fold

Figure 4.1 (Branden & Tooze). Alpha\beta domains are found in many proteins. They occur in different classes, two of which are shown here. **(a)** a closed barrel exemplified by schematic and topological diagrams of the enzyme triosephosphate isomerase. **(b)** An open twisted sheet with helices on both sides, as in the coenzyme-binding domain of some dehydrogenases. Both classes are built up from β–α–β motifs that are linked such that the β strands are parallel. Rectangles represent α helices, and arrows represent β strands in the topological diagrams.

a) This contains an open twisted β sheet surrounded by α helices.
b) A typical example is the nucleotide (NAD^+)-binding domain of enzyme lactate dehydrogenase.
c) This domain is sometimes called the **Rossman fold** after its discoverer Michael Rossman of Purdue University.

iii) Leucine-rich motifs (or horseshoe fold)

a) This class is formed by amino acid sequences that contains repetitive regions of a specific pattern of leucine residues, so-called **leucine-rich motifs**, which forms α helices and β strands.
b) The β strands form a curved parallel β sheet with all the α helices on the outer side.
c) The structure of one member of this class, a ribonuclease inhibitor, is shaped like a horseshoe, and consequently this classis also called the **horseshoe fold**.

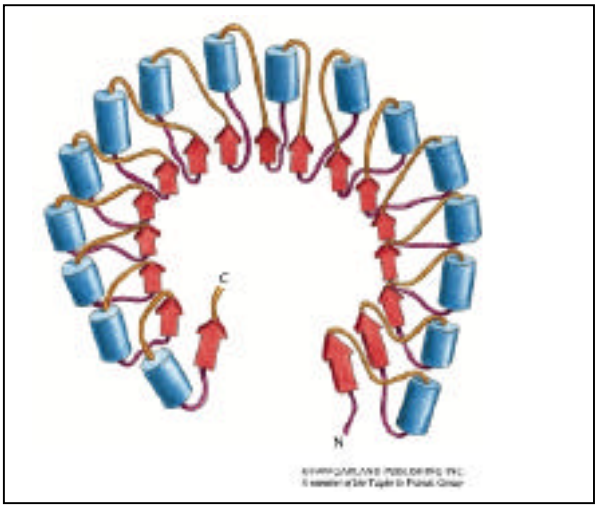


Figure 4.11 (Branden & Tooze). Schematic diagram of the structure of the ribonuclease inhibitor. The molecule, which is built by repetitive β-loop-α motifs, resembles a horseshoe with a 17-stranded parallel β sheets on the inside and 16 α helices on the outside.

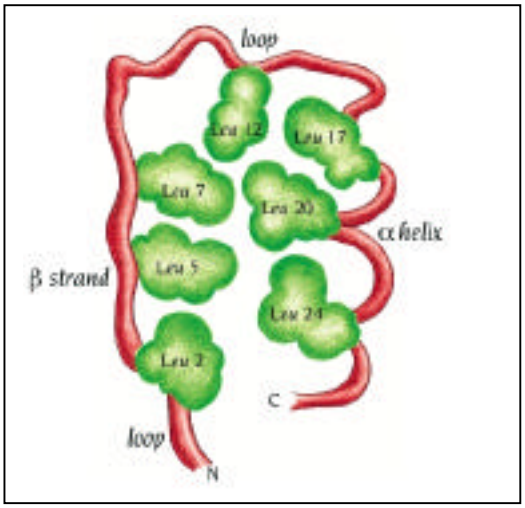


Figure 4.12 (Branden & Tooze). Schematic diagram illustrating the role of the conserved leucine residues, in the leucine-rich motif in stabilizing the β-loop-α structural module.

PROTEIN STRUCTURE I & II

Note:
Four types of interactions cooperate in stabilizing tertiary structures of globular proteins, they are **disulfide bonds** (will be discussed), **hydrophobic interaction**, **hydrogen bonds** and **ionic interactions**.

IV) QUATERNARY STRUCTURE

1) Protein molecules that have only one chain are called monomeric proteins, but a fairly large number of proteins have **quaternary structure**, which consists of several identical polypeptide chains (**subunits**) that associate into a **multimeric** molecule in a specific way. Different polypeptide chains can be called subunits, monomers or protomers.
2) Each polypeptide-chain subunit is usually folded into an apparently independent globular conformation, which then interacts with other monomers.
3) These subunits can function either **independently** of each other or **cooperatively** so that the functional state of one subunit is dependent on the functional state of other subunits.
4) Some protein molecules are assembled from several **different subunits** with **different functions**, for example RNA polymerase II.
5) Quaternary structure is invariably observed in the crystal structure of the protein.

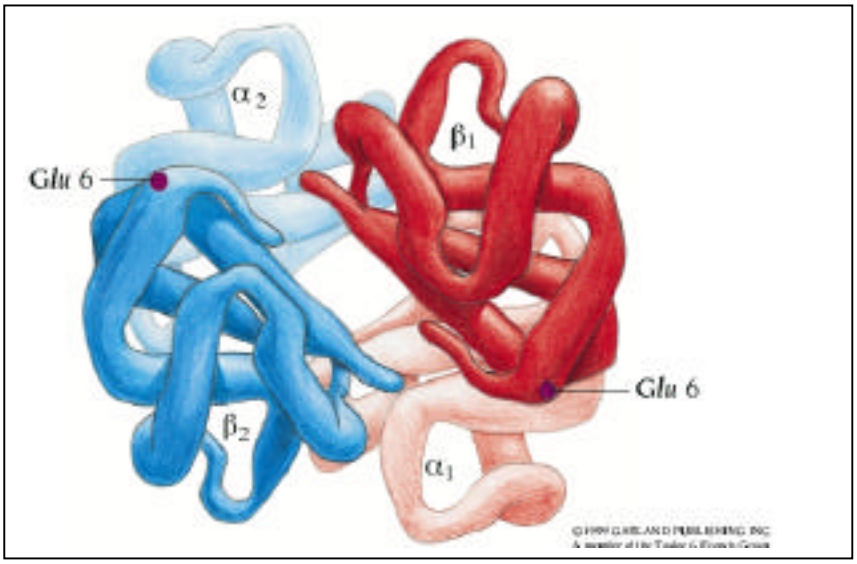


Figure 3.13 (Branden & Tooze). The hemoglobin molecule is built up of four polypeptide chains: two α chains and two β chains. Each chain has a three-dimensional structure similar to that of myoglobin: the globin fold.

PROTEIN STRUCTURE I & II

Note: Subunits are held together by noncovalent interactions (**hydrophobic interactions, hydrogen bonds, ionic bonds**) as in hemoglobin.

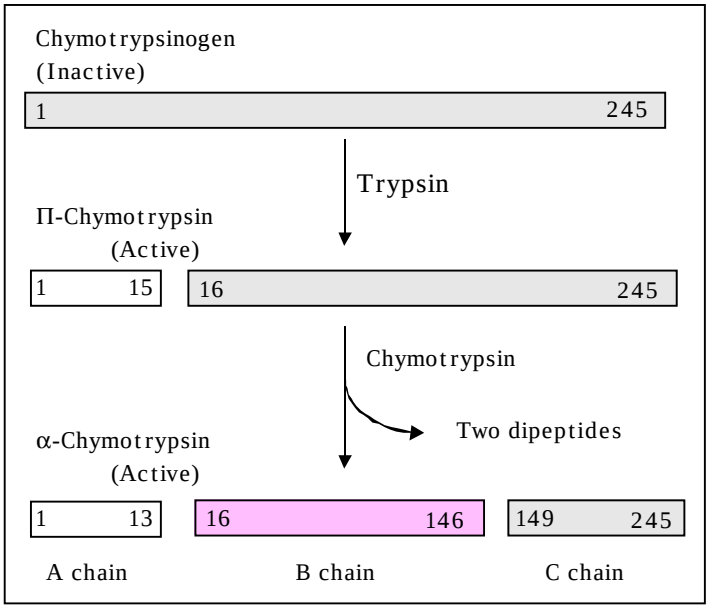
POST TRANSLATIONAL MODIFICATION OF PROTEIN

a) Many of the 20 amino acids can be modified to enhance structural and functional capabilities of protein.
b) Release of the completed polypeptide chain from the ribosome is often not the final step in formation of a protein. Various covalent modifications often take place, either during or after assembly of the polypeptide chain.
c) Because the list of modification is constantly growing, only those that occur more frequently or have the most important biological effect will be discussed.

Proteolytic processing:

a) Proteolytic cleavage of polypeptide chains after synthesis is a common occurrence in proteins destined for cellular organelles or for **secretion**. This includes removal of the **signal peptide**.

b) Proteolytic enzymes in the digestive tract (such as chymotrypsin) are produced in inactive forms that are generally designated as **zymogens** (or proenzymes) which undergo proteolytic cleavage for activation.
c) Proteolytic cleavage of a **proenzyme** may initiate a series of local conformational rearrangements within the protein molecule resulting in creation of the active site.



Glycosylation:

a) Many **secretory proteins** and **membrane proteins** bear covalently attached carbohydrate moieties.

PROTEIN STRUCTURE I & II

b) The oligosaccharide units of glycoproteins are linked to either **aspargine** side chains by **N-glycosydic** bonds or to **serine** and **threonine** side chains by **O-glycosydic** bonds.

c) N-linked glycosylation may help recognition of the protein molecules by chaperones (such as calnexin) and thus ensure proper folding of the polypeptides.

d) Most proteins, such as antibodies, that are secreted by cells acquire **carbohydrate** units on specific asparagine residues.

e) The addition of sugars makes a protein more hydrophilic.

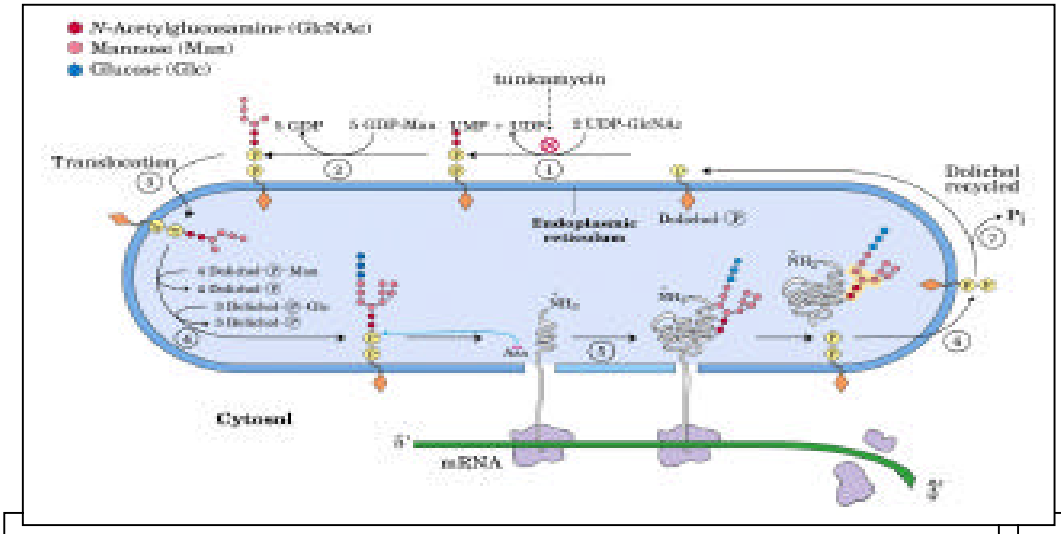


Figure 27-36 (Nelson and Cox). Synthesis of the core oligosaccharide of glycoproteins.

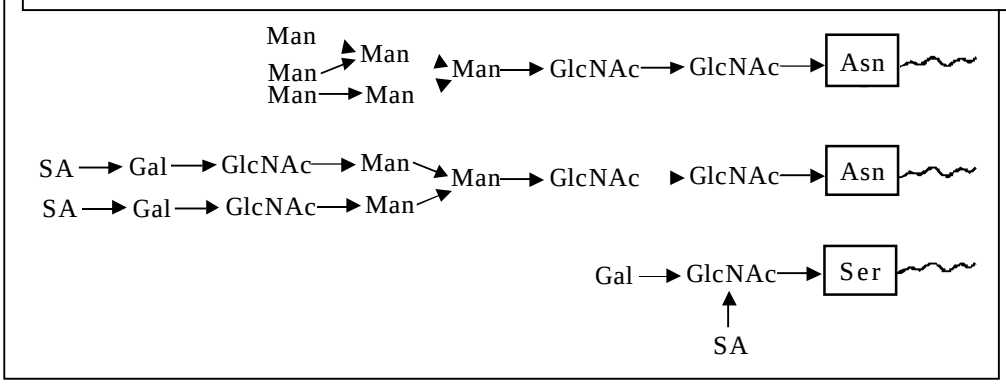
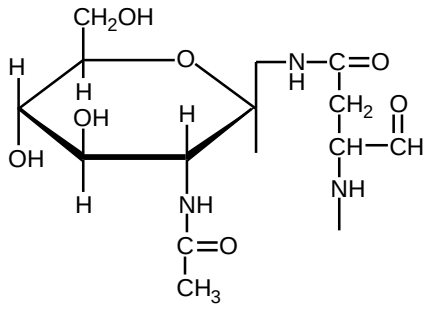


Figure 2.21 (Creighton). Structures of typical polysaccharide chains attached to proteins.



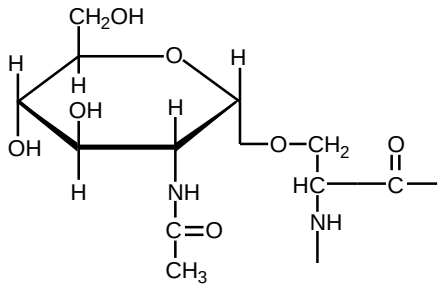
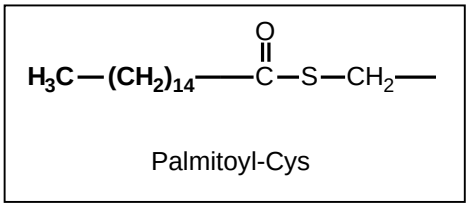


Figure 2.20 (Creighton). Structures of carbohydrate-protein linkages commonly found in proteins.

PROTEIN STRUCTURE I & II

a) Lipids frequently tether intrinsicallysoluble protein to membranes.
b) The polar groups of the lipid is attached covalently to the protein, while the hydrophobic portion of the lipid is embedded in the membrane.
c) Lipids, such as **Myristoyl** groups can be attach to the N-terminus and **glycosyl-phosphatidylinositol** and **farnesyl** groups to the C-terminus of the protein molecules.



Palmitoyl-Cys

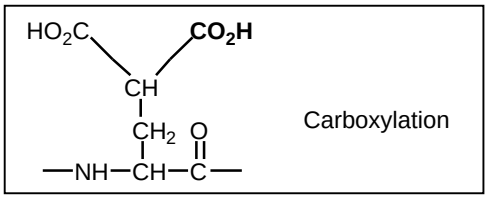
d) **Palmitoyl** groups can be attached to the side chains of Cys residues. Modified proteins are usually firmly anchored to the membrane by the palmitoyl group.

Sulfation:

a) Sulfation of **tyrosine** residue is another posttranslational modification that is limited to protein that passes through the Golgi apparatus.
b) The functional roles of sulfation are just being discovered.
c) There are indications that sulfation affects the biological activities of neuropeptides, the proteolytic processing of some protein precursors, and intracellular transport of some secretory proteins.

Carboxylation:

a) Certain **glutamate residues**, particularly in proteins that involved in blood clotting and bone structure, are carboxylated to yield the unusual residue **γ-carboxyglutamic acid (Gla)**.
b) The function of Gla residues is almost invariably linked to **binding of Ca^{2+} ions**. The second adjacent carboxyl group considerably increases the intrinsic ability of these residues to bind calcium ion.

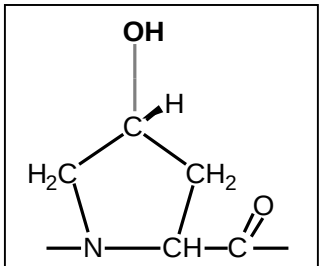


Carboxylation

c) The modification is necessary for the functional properties of the protein.

Hydroxylation:

a) Hydroxylation of certain proline (**Pro**) and lysine (**Lys**) residues is an important step in maturation and secretion of collagen.
b) Pro residues that occur in the sequence **-Xaa-Pro-Gly-** are hydroxylated on the **γ-carbon**, and Lys residues in the sequence **–Xaa-Lys-Gly-** are hydroxylated on the **δ-carbon**.These modifications are vital for the folding and



Hydroxylation

assembly of **mature collagen**. The biological significance of this modification is evident in **scurvy**, which results from insufficient hydroxylation of collagen because of a deficiency of vitamin C.

d) A similar hydroxylation has recently been found to occur on certain Aspargine (**Asn**) and Aspartate (**Asp**) residues in a few proteins.

e) These modifications might have some role in **Ca^{2+} binding**.

Phosphorylation:

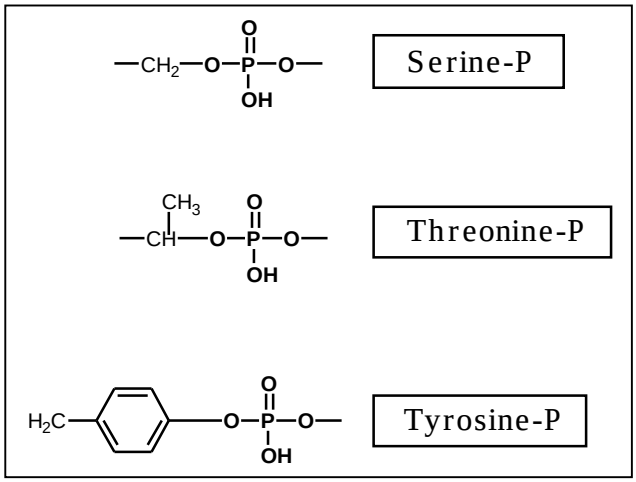
a) A large number of proteins are phosphorylated at specific sites, usually reversibly and with important functional consequences.

b) The sites for phosphorylation are usually the **hydroxyl groups** of **Ser**, **Thr**, or **Tyr** residues, but **Asp**, **His**, and **Lys** residues may also be phosphorylated.

c) The phosphoryl groups are added, by using specific protein **kinases**, using **ATP** as the phosphoryl donor. The phosphate groups on these three modified amino acids can readily be removed, enabling them to act as **reversible switches** in regulating cellular process.Reversible phosphorylation of proteins is frequently one step in the action of **hormones**.

d) From the point of view of protein structure and function, the most important aspect of the phosphoryl group appears to be its **negative charge**.

e) Not all phosphorylation is functionally important. Phosphorylation of milk protein casein is probably primarily of nutritional importance.



Acetylation:

a) Acetylation of the amino terminal end of many proteins makes them more **resistant to degradation**.

b) A variety of **N-acetyltransferase** enzymes are thought to catalyze this reaction, using acetyl-CoA as the acetyl donor.

c) Acetylation of amino acid residues within the DNA-binding domains of certain transcription factor (such as p53 and E2F1) **alters DNA-binding properties** of those proteins.

d) Acetylation also regulates **protein-protein interaction**, as in *Drosophila* transcription factor TCF.

PROTEIN STRUCTURE I & II

Cysteines and disulfide bonds:

a) Two cysteine residues in different parts of the polypeptide chain but adjacent in the three-dimensional structure of a protein can be **oxidized** to form a **disulfide bridge**.

b) This reaction requires an **oxidative environment**, and such disulfide bridges are usually not found in intracellular proteins, which spend their lifetime in an essentially reductive environment. Disulfide bridges **stabilize three-dimensional structure**. In some proteins these bridges hold together **different polypeptide chains**, for example the A- and B- chains of insulin are linked together by two disulfide bridges.

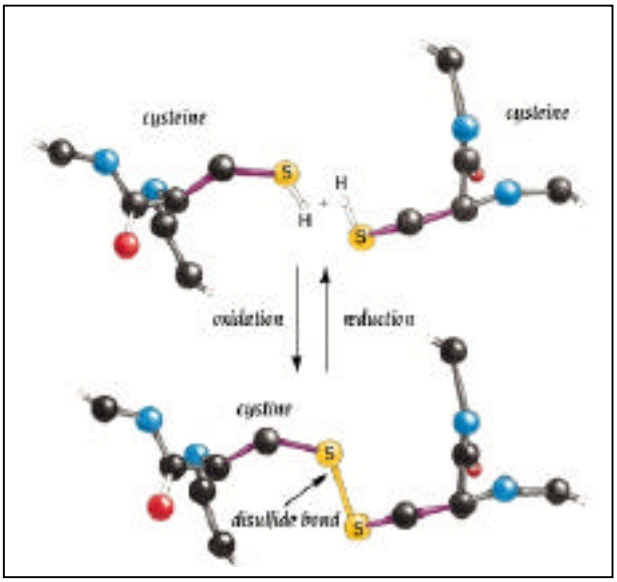


Figure 1-4 (Branden and Tooze). A disulfide bridge (-S-S-) is formed from the sulfhydryl groups (-SH) of two cysteine residues. The product is a cystine residue.

c) In **bacteria** oxidation of sulfhydryl groups of cystein residues occurs mainly in the **periplasmic space,** while in eukaryotic cells it occurs in the **endoplasmic reticulum**.

Isomerization of proline residues:

a) Peptides bonds could either be of cis- or trans- conformation. The so-called **cis-peptide** conformation is the less stable form of peptide group and are rarely occurs in proteins.

b) However, **cis-proline peptides** may occur in some proteins, generally in the **tight bends** of polypeptide chain and are sometimes important for activity or for conformational flexibility.

c) In the native protein these less stable cis-proline peptides are stabilized by the tertiary structure.

d) **Cis-trans isomerization** of proline peptides is intrinsically a slow process and, in vitro,

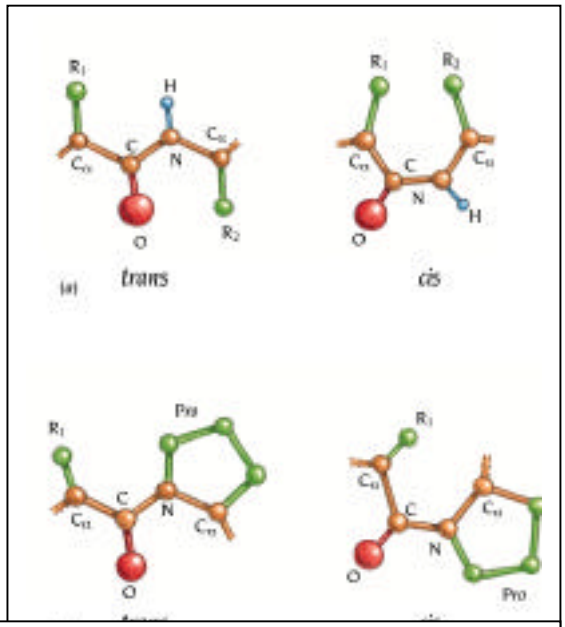


Figure 6.9 (Branden & Tooze). (a) Peptide can adopt two different conformations, *trans* and *cis*. In the trans-form the C=O and the NH groups point in opposite direction. For most peptides the *trans*-form is about 1000 times more stable than the *cis*-form. **(b)** When the second residue in a peptide is proline the *trans*-form is only about four times more stable than the cis-form. Cis-proline peptides are found in many proteins.

it is frequently the rate-limiting step.In vivo, the rate of isomerization is enhanced by the enzymatic activity of **peptidyl prolyl isomerases**, which are found in both prokaryotic and eukaryotic organisms.

MANY PROTEINS CONTAIN METAL ATOMS

a) Although the side chains of 20 different amino acids have very different chemical properties, their **chemical versatility is not unlimited**, and for some functions of proteins metal ions are exploited for suitable and more efficient reaction.

b) Side chains of **histidine, cysteine, aspartic acid**, and **glutamic acids** are excellent **metal ligands**, and a fairly large number of proteins have recruited metal atoms as intrinsic parts of their structures.

c) Among the most frequently used metals are **iron, zinc, magnesium** and **calcium**.

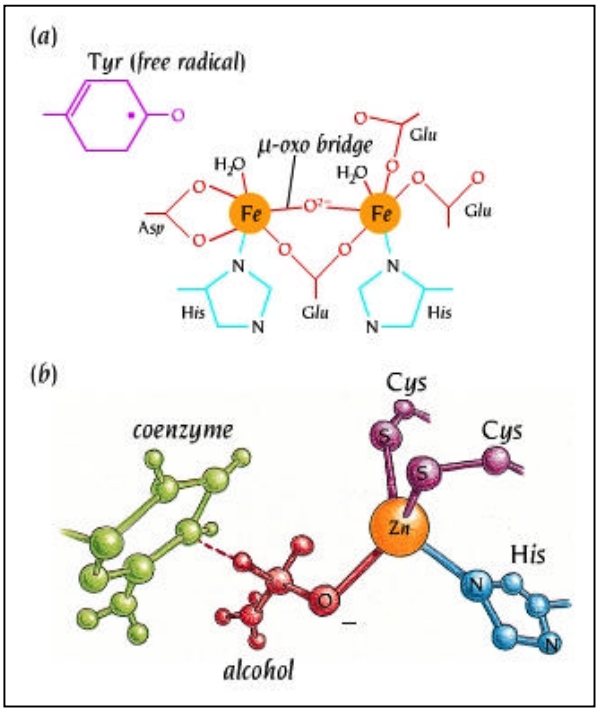


Figure 1.9 (Branden and Tooze). Examples of functionally important intrinsic metal atoms in proteins. (a) The di-iron center of the enzyme ribonuclotide reductase. Two iron atoms form a redox center that produces a free radical in a nearby tyrosine side chain. The coordination of the iron atom is completed by histidine, aspartic acid, and glutamic acid side chains as well as water molecules. (b) The catalytically active zinc atom in the enzyme alcohol dehydrogenase. The zinc atom is coordinated to the protein by one histidine and two cysteine side chains. During catalysis zinc binds an alcohol molecule in a suitable position for hydride transfer to the coenzyme moiety, a nicotinamide.

PROTEIN STRUCTURE I & II

Examples: A) **Hemoglobin** uses iron for carrying oxygen molecule. B) Zinc is used to stabilize the DNA-binding regions (motifs) of a class of transcription factors called **zinc fingers**.

STRUCTURE-FUNCTION RELATIONSHIP

Proteins as a class of molecules are unique in being able to recognize and interact with highly diverse molecules. Proteins are able to interact specifically with such a wide range of structures because they are highly proficient at forming **complementary surfaces and clefts**. The rich repertoire of side chains on these surfaces and these clefts enables proteins to form **covalent and noncovalent interactions.**

MYOGLOBIN AND HEMOGLOBIN: BINDING AND TRANSPORT OF SMALL MOLECULE

Myoglobin and hemoglobin are the two most-studied and best-understood proteins. The subunits of hemoglobin closely resemble myoglobin in their three-dimensional structure. Oxygen is poorly soluble in aqueous solutions and cannot be carried to tissues in sufficient quantity if it is simply dissolved in blood serum. The function of shuttling oxygen molecules in blood and in muscle tissues is accomplished by the action of hemoglobin and myoglobin, respectively.

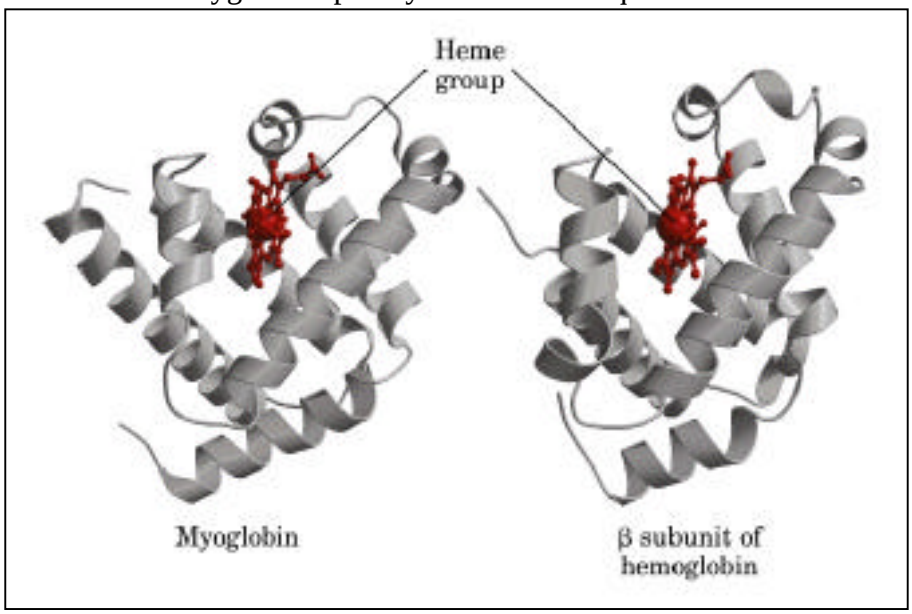


Figure 7-6 (Nelson and Cox). Comparison of the conformations of the main chain of myoglobin and the β chain of hemoglobin. The similarity of their conformations is evident.

PROTEIN STRUCTURE I & II

Myoglobin

a) Myoglobin, the **oxygen carrier** in muscle, is a single polypeptide chain of 153 amino acids.

b) The capacity of myoglobin to reversibly bind oxygen depends on the presence of **heme**, a nonpolypeptide **prosthetic (helper) group** consisting of **protoporphyrin** and a central **iron atom**.

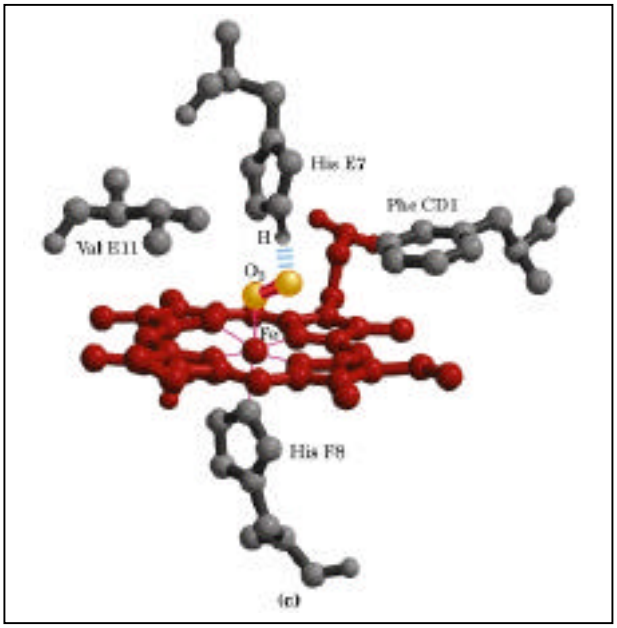


Figure 7-5 (Nelson and Cox). O_2 bound to heme group of myoglobin. The iron atom of heme group is firmly liganded on its proximal side chain to a His (His F8). The O_2 molecule is bound on the distal side to the iron atom and is also hydrogen-bonded to a distal His (His E7).

c) Myoglobin is an extremely **compact** molecule and is built primarily of α **helices** and **turns** between the helices.

d) The **interior** of myoglobin consists almost entirely of **nonpolar** residues, such as leucine, valine, methionine, and phenylalanine. The only polar residues inside are **two histidines**, which play critical roles in the **binding of heme oxygen.**

d) The interaction of O_2 with heme is greatly affected by protein structure and is often accompanied by **conformational changes**.

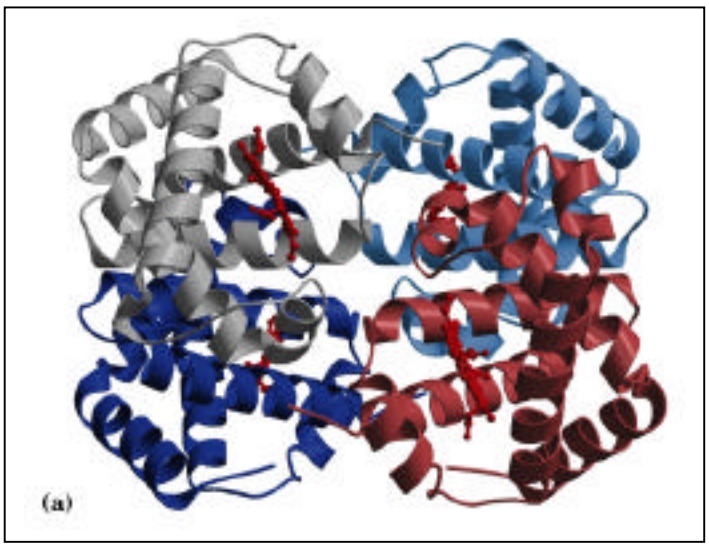
e) The heme molecule is **deeply buried** in the folded polypeptide, with no direct path for oxygen to enter ligand-binding site from the surrounding solution.

f) Rapid molecular flexing of the amino acid side chains produces transient cavities in the protein structure, and O_2 evidently makes its way in and out moving through these cavities.

Hemoglobin is a tetramer of myoglobin-like subunits

i) Hemoglobin, in contrast to myoglobin, is a tetrameric protein and contains four heme prosthetic groups, one associated with each polypeptide chain.

j) Adult hemoglobin contains two types of globin, two α chains (141 residues each) and two β chains (146 residues each).



k) Although a fewer than half of the amino acid residues in the polypeptide sequences of the α and β subunits are identical, the three-dimensional structures of the two types of subunits are very similar.

l) The quaternary structure of hemoglobin features strong interactions between unlike subunits. Hydrophobic interaction predominates at the interfaces, but there are also many hydrogen bonds and a few ion pairs (some referred to as salt bridges).

Figure 6-23 (Nelson and Cox). Quaternary structure of hemoglobin

m) The binding of O_2 to a hemoglobin subunit triggers a **change in conformation**. When the entire protein undergoes this transition the structures of the individual subunits change very little. In this process, some of the ion pairs are broken and some new ones are formed.

STRUCTURAL ORGANIZATION OF COLLAGEN: MECHANICAL SUPPORT

a) Collagens are proteins that assemble into **fibrous supermolecular aggregates** in the extracellular space.

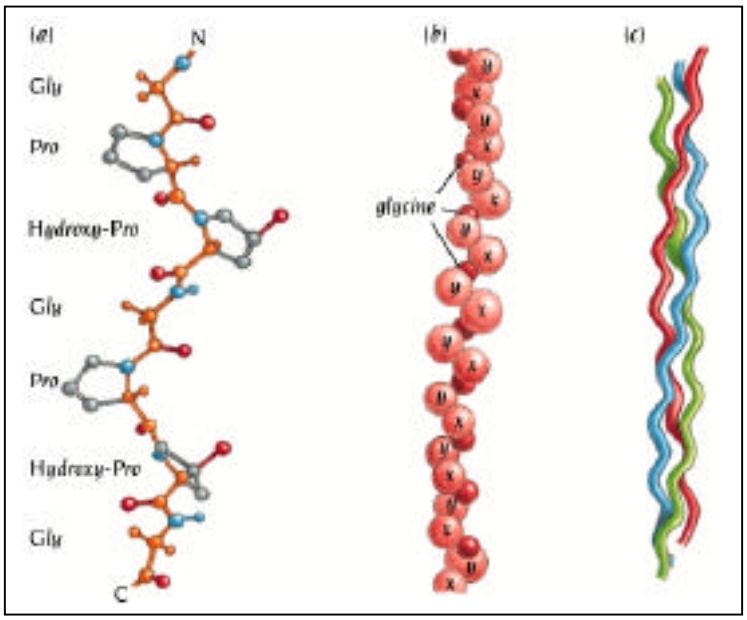
b) Collagen is a **right-handed superhelix** formed by three parallel, very extended left-handed helices. The **three polypeptide chains** contains a large number of repeat sequences **Gly-X-Y** where "X" is often **proline** and "Y" is often **hydroxyproline.**

c) Each collagen polypeptide chain contains about **1000 amino acid** residues and the entire triple chain is about **3000Å long**.

d) In this superhelix the side chain of every third residue is close to the central axis where there is no space available for a side chain, consequently **every third residue is a glycine**. Any other residue deforms the superhelix and certain inherited connective tissue diseases are due to mutations in the codon for these glycine residues.

Figure 14.1 (Branden & Tooze). Each polypeptide chain in the collagen molecule folds into an extended polyproline type II helix with a rise per turn along the helix of 9.6 Å comprising 3.3 residues. In collagen molecule three such chains are supercoiled about a common axis to form a 3000-Å-long rod-like molecule. **(a)** Ball and stick model of two turns of a polypeptide chain. **(b)** A model of one collagen chain in which each amino acid is represented by a sphere **(c)** A small part of a collagen superhelix showing all three chains.

PROTEIN STRUCTURE I & II



e) In the regular triple helix collagen structure the three chains are held together by direct **hydrogen bonds** between proline C=O groups of one chain and the glycine NH groups of another.

f) All side chains as well as the C=O groups of glycines in all three chains are on the outside of the triple helix molecule and in contact with **water molecules**. This water molecules mediate hydrogen bonds between the hydroxyl groups of hydroxyproline and the peptide C=O and NH groups both within each chain and between different chains. These **water-mediated hydrogen bonds** are essential for the stability of the triple helix and are presumable the reason for the presence of hyroxyproline in collagen.

OSTEOGENESIS IMPERFECTA:

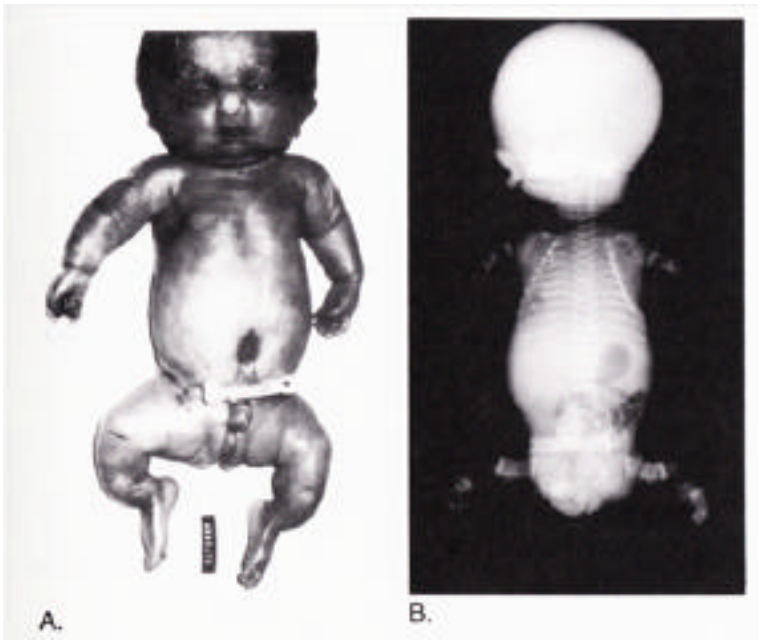
Collagen has evolved to provide strength and is found in connective tissues such as tendons, cartilage, the organic matrix of bone, and the cornea of the eye. The unusual amino acid content of collagen helix is related to structural constraints unique in collagen helix. Some human genetic defect in collagen structure illustrates the close relationship between amino acid sequence and three-dimensional structure in this protein. **Osteogenesis imperfecta (OI type II)** is characterized by abnormal bone formation in babies. It can be lethal and result from substitution of an amino acid residue with a larger R group (such as Cys or Ser) for a single Gly residue in each α-chain. These single-residue substitutions have a catastrophic effect on collagen function because they disrupt the Gly-X-HyPro repeat that gives collagen its unique helical structure.

Disease	Clinical Features	Inheritence	Biochemical Effects
OI Type II	Lethal in perinatal period, minimal calvarial mineralization, beaded ribs, compressed femurs, marked long bone deformity, platyspondyly	Autosomal Dominant (new mutation)	Rearrangements in the COLA1 and COLA2 genes; substitutions for glycyl residues in the triple-helical domain of the α1(I) or α2(I) chain; smalldeletion in α2(I) on the background of a null

PROTEIN STRUCTURE I & II

			allele

Fig. 137-4. Clinical (A) and radiographic (B) appearance of an infant with perinatal lethal form of osteogenesis imperfecta (OI Type-II)



(The Metabolic and Molecular Basis of Inherited Disease, Vol. III, Seventh Edition)

ZINC FINGERS: DNA BINDING

a) More than a thousand different **transcription factors** contain zinc as an essential element of their DNA-binding domains.
b) The polypeptide chains of such zinc-containing motifs are about **50 amino acids or less**, with regular patterns of **cysteines** and/or **histidine** residues along the chains. These residues bind to **zinc atoms**, thereby providing a scaffold for the folding of the motif into a small compact domain.
c) The zinc atom is buried in the interior of the protein and is necessary for the formation of a stable **finger structure**.
d) The residues between the second cysteine zinc ligand and the first histidine ligand, of the classic zinc finger motif, form the "**finger region**". This is the region of the polypeptide chain that forms the **main interaction area** with **DNA**.

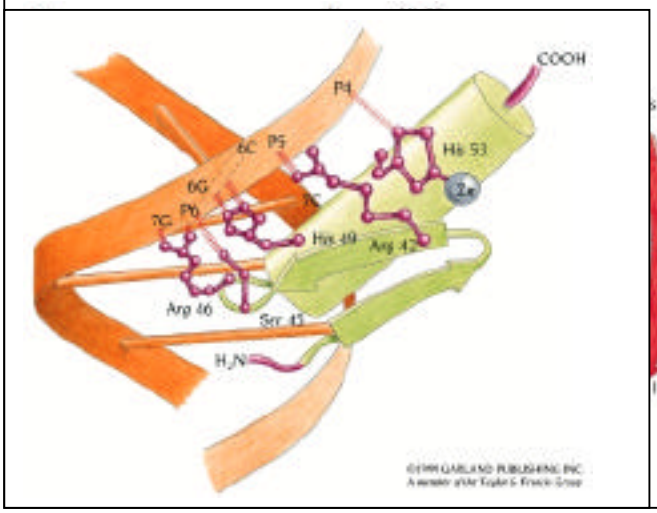


Figure 10.1 (Branden & Tooze). (a) The classic Zinc finger motif comprises about 30 amino acid residues, with two

Figure 10.4 (Branden & Tooze). Detailed view of the binding of the second zinc finger of a protein called Zif 268 to DNA. Two side chains, Arg 46 and His 49, form sequence-specific interactions with DNA. There are also three nonspecific interactions between phosphate groups of the DNA and the side chains of Arg 42, Ser 45 and His 53.

PROTEIN STRUCTURE I & II

TESTICULAR FEMINIZATION:

Steroid hormones mediate their action through binding to their specific nuclear receptors. The hormone-receptor complex acts as a transcriptional activator by binding to DNA sequences known as hormone-response elements. These receptors have distinct functional domains for hormone and DNA binding. DNA binding, is mediated through zinc figure motifs.

Male sex differentiation and development proceed under direct control of male sex hormones testosterone and 5α-dihydrotestosterone (known as **androgens**). The actions of both the steroids are mediated by **androgen receptor** (AR), which has two functionally distinct DNA-binding zinc fingers. Single amino acid substitutions (557C→Wand 574C→R) (Fig. 95-15) have been reported in the zinc finger regions of human androgen receptor in individuals with "**Complete"** or **"Incomplete Testicular Feminization"** (Fig. 95-11). The mutant receptors of this type have little or no effect on hormone binding but impair binding to DNA and are thus incapable of mediating transcriptional activation. The clinical features of complete testicular feminization are summarized in the following table (Form "The metabolic and Molecular Basis of Inherited Disease" seventh edition, McGraw-Hill, Inc.). This is a classic example of how disruption of protein structural motifs may lead to serious physiological abnormalities.

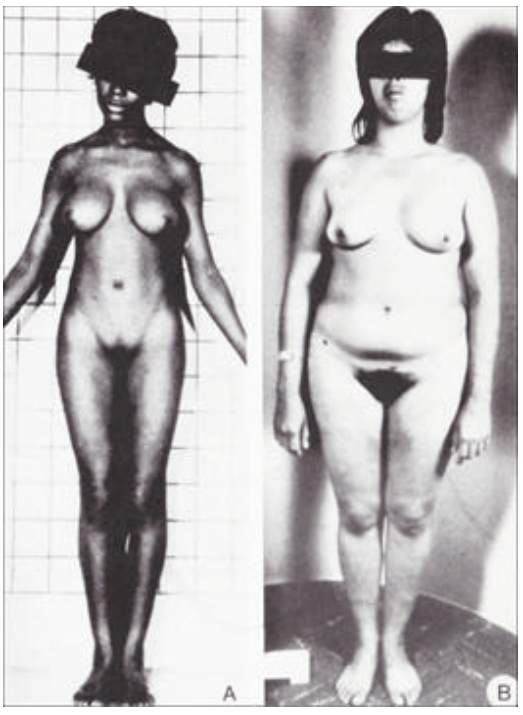


Figure 95-11. Two patients with disorders of androgen receptor function. (A) Complete testicular feminization. (B) Incomplete testicular feminization (The metabolic and Molecular Basis of Inherited Disease" seventh edition, McGraw-Hill, Inc.).

(The Metabolic and Molecular Basis of Inherited Disease, Vol. III, Seventh Edition)

PROTEIN STRUCTURE I & II

LEUCINE ZIPPER:PROTEIN-PROTEIN INTERACTION

a) Leucine zippers provide dimerization interactions for some eukaryotic transcription factors, such as GCN4, C/EBP, Fos, Jun, and Myc.

b) These proteins have regions of about 30 amino acid residues, with sequences arranged in **modules of 7 residues (heptad repeats)**.

c) The fourth residue of each module is almost always a **leucine**, hence the name **leucine zipper**. In addition, the first residue of each module is frequently hydrophobic.

d) Such modules **dimerize** and forms **two parallel coiled-coil** α **helices** with a helical repeat of 3.5 residues per turn.

The hydrophobic side chains of the **heptad repeats** form a hydrophobic core between the helices in this coiled-coil with leucine residues facing each other. This hydrophobic core is one major determinant of the stability of the dimer.

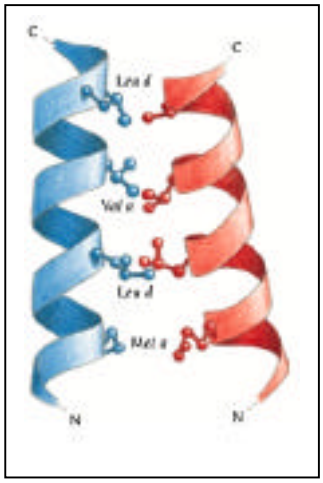


Figure 10.18 (Branden & Tooze). Side-chain interactions in the leucine zipper structure. The hydrophobic side chains in spikes a and d form a hydrophobic core between the two coiled α helices.

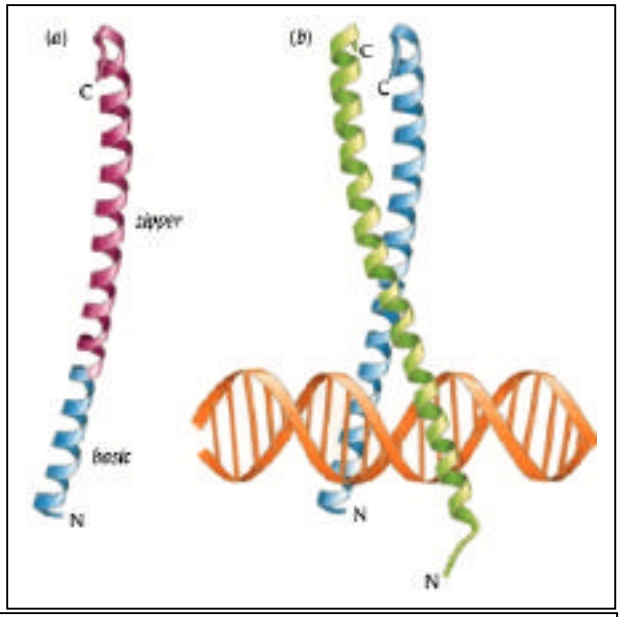


Figure 10.21 (Branden & Tooze). The structure of a complex between the DNA-binding domain of GCN4 and a fragment of DNA. **(a)** Each monomer of the GCN4 domain forms a smoothly curved continuous α helix comprising both the basic and the leucine zipper regions. **(b)** The monomers are held together in a dimer in the zipper region.

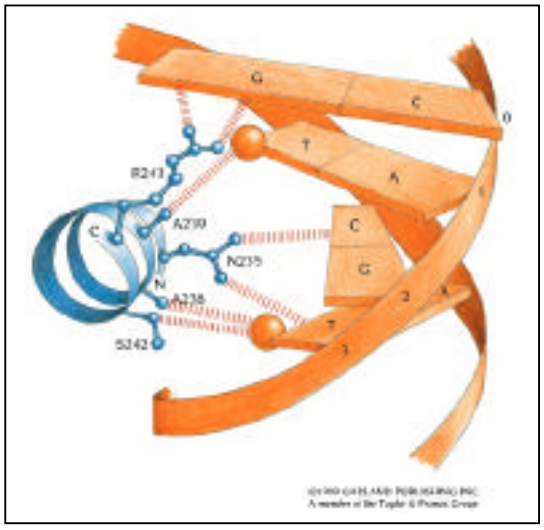


Figure 10.22 (Branden & Tooze). Sequence-specific interactions between one of the α-helical basic regions of GCN4 and bases in the DNA fragment. The methyl group thymine bases are shown as spheres.

PROTEIN STRUCTURE I & II

PROTEIN FOLDING

The process by which polypeptide chain acquires its correct three-dimensional structure to achieve the biologically active **native state** is called protein folding. Although some polypeptide chains **spontaneously** fold into the native state, other requires the assistance of enzymes, for example, to catalyze the formation and exchange of **disulfide bonds**; and many requires the assistance of a class of proteins called **chaperones**. A chaperone binds to a polypeptide chain and prevents it from making illicit associations with other folded or partly folded proteins, hence the name chaperone. A chaperone also promotes the folding of the polypeptide chain it holds.

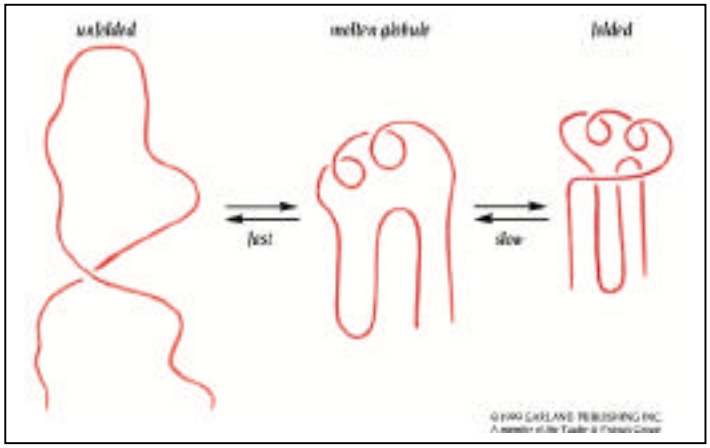


Figure 6.2 (Branden and Tooze). The molten globule state is an important intermediate in the folding pathway when a polypeptide chain converts from an unfolded state. The molten globule has most of the secondary structure of the native state but it is less compact and the proper packing interactions in the interior of the protein have been formed.

After a polypeptide has acquired most of its correct secondary structure, with the α helices and the β sheets formed, it has a looser tertiary structure than the native state and is said to be in the **molten globular state**. The compaction that is necessary to go from the molten globular state to the final native state occurs spontaneously. The process of protein folding is largely dictated by the amino acid sequence of the polypeptide.

RIBONUCLEASE: THE AMINO ACID SEQUENCE OF A PROTEIN SPECIFIES ITS THREE-DIMENSIONAL STRUCTURE

The fact that the **amino acid sequence** of a protein contains the information for its folding into an active **conformation** was first recognized in ribonuclease.

1) Ribonuclease A is a **pancreatic enzyme** that hydrolyzes RNA molecules.
2) It is consist of a single polypeptide chain of 124 residues that fold mainly into β **strands** which forms a tightly packed hydrophobic interior.
3) The enzyme is further stabilized by formation of four **disulfide bonds**, which can be cleaved reversibly by reducing with a reducing agent **β-mercaptoethanol**.

4) In the presence of a large excess of β-mercaptoethanol disulfides are reduced, so that the final product is a protein in which the disulfides are fully converted in **sulhydryls** (cysteines).
5) Treatment of ribonuclease with β-mercaptoethanol in the presence of 8 M urea (a denaturing agent) results in a fully reduced, randomly coiled polypeptide chain devoid of any enzymatic activity.
6) Denatured ribonuclease, freed of urea and β-mercaptoethanol by **dialysis**, slowly **regains enzymatic activity**.
7) The sulfhydryls of the denatured enzyme become oxidized in contact with air, and the enzyme **spontaneously refolded** into a catalytically active form.
8) Therefore, **information needed** to specify the complex **three-dimensional structure** of ribonuclease is contained in its amino acid sequence.

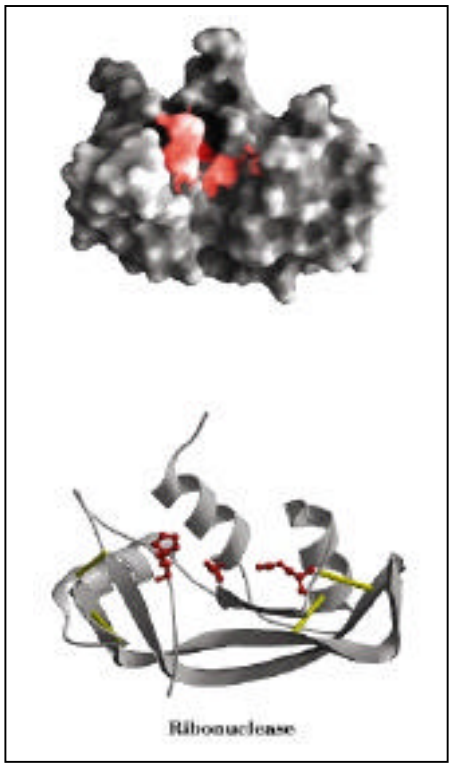


Figure 6-18 (Nelson and Cox). Three dimensional structure of Ribonuclease shown in surface contour (top) and ribbon representations (bottom). There are four disulfide bonds in ribonuclease.

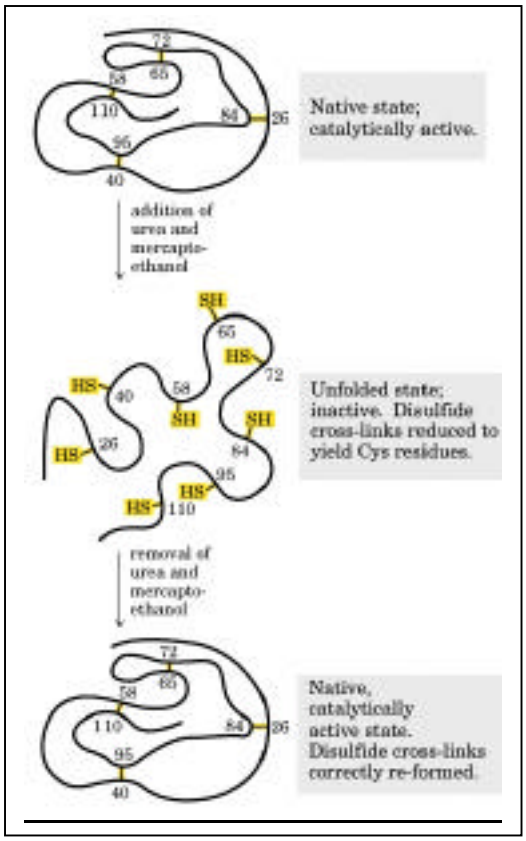


Figure 6-27 (Nelson and Cox). Renaturation of unfolded, denatured ribonuclease. Urea is used to denature ribonuclease, and mercaptoethanol ($HOCH_2CH_2SH$) to reduce and thus cleave the disulfide bonds to yield eight Cys residues. Renaturation ivolves restablishment of the correct disulfide cross-links.

PROTEIN STRUCTURE I & II

CYSTIC FIBROSIS:

Cystic fibrosis (CF) is a serious and relatively common hereditary disease of humans. Presence of two defective copies of genes cause severe symptoms of the disease, which include obstruction of the gastrointestinal and the respiratory tracts, commonly leading to severe bacterial infection of the airways and death due to respiratory insufficiency before the age of 30. In CF, the thin layer of mucus that normally coats the internal surfaces of the lungs is abnormally thick, obstructing air flow and providing a haven for pathogenic bacteria, particularly *Staphylococcus aureus* and *Pseudomonas aeruginosa*.

The defective gene in CF patients was discovered in 1989. It encodes a membrane protein called Cystic Fibrosis Transmembrane conductance Regulator, or **CFTR**. CFTR is an ion channel specific for Cl^-. The mutation responsible for CF in 70% of cases results in deletion of a **Phe residue at position 508**, with the effect that the mutant protein is not correctly folded and inserted in the plasma membrane. The result is a nonfunctional Cl^- channel in the epithelial cells that line the airways, the digestive tract, and exocrine glands.

Normally, epithelial cells that line the inner surface of the lungs secrete a substance that traps and kills bacteria, and the cilia on the epithelial cells constantly sweep away the resulting debris. According to one hypothesis, the bactericidal activity requires a relatively low NaCl, concentration. In CF patients, the defect in Cl^- channel of CFTR results in surface fluid containing a high concentration of NaCl, and this fluid is much less effective in killing bacteria. Frequent infections by bacteria such as *S. aureus* and *P. aeruginosa* progressively damage the lungs and reduce respiratory efficiency.

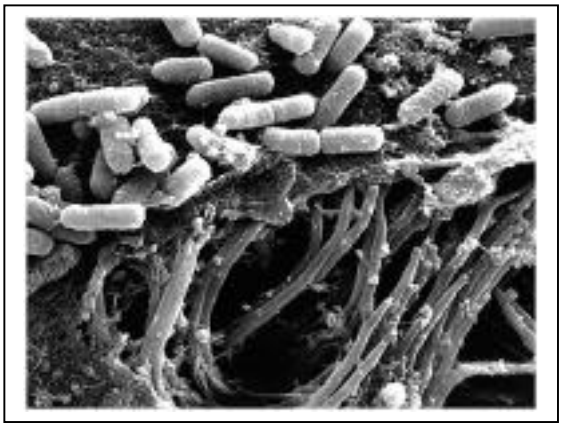


Figure 1 (Lehninger). Mucus lining the surface of the lungs traps bacteria, In healthy lungs, these bacteria are killed by the action of cilia. In CF, the bactericidal activity is impaired, resulting in recurring infections and progressive damage to the lungs.

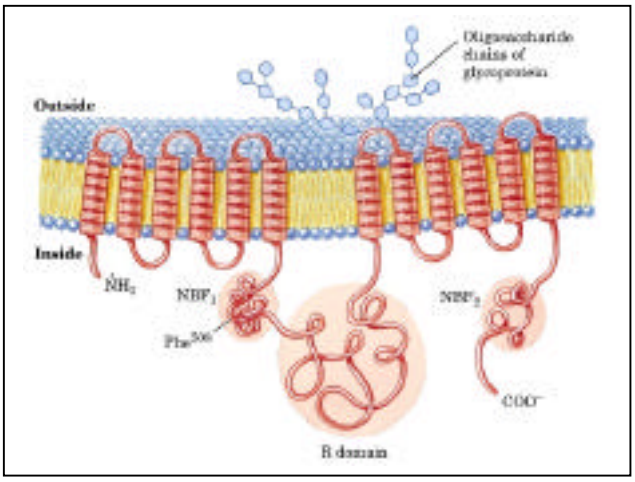


Figure 2 (Lehninger). The cystic fibrosis transmembrane conductance regulator, CFTR. The most commonly occurring mutation leading to CF is the deletion of Phe^{508}, in the NBF1 domain.

PROTEIN STRUCTURE I & II

PHYSICAL BASIS OF PROTEIN DENATURATION

A loss of three-dimensional structure sufficient to cause loss of function is called denaturation.

Extremes of pH:

a) Many proteins unfold at pH values less than about 5 and greater than 10.
b) Unfolding at such extreme pH generally occurs because the folded proteins have groups buried in **nonionized** form that can ionize only after unfolding. Most prevalent are **Met** and **Tyr** residues, which tend to cause unfolding at acid and alkaline pH values, respectively.
c) **Salt bridges** between ionizing groups, which contribute to the structural stability of the protein, can be disrupted by extreme pH values, in which one of the interacting groups is no longer ionized.

Effect of Denaturants:

a) A large number of reagents affect protein stability when added to the aqueous solvent. Those that decrease protein stability are known as denaturants, such as **urea** and **guanidinium** chloride.
b) They act by **breaking inherent hydrogen bonds** and by interacting with the peptide groups in unfolded proteins by hydrogen bonding.
c) Both urea and guanidinium chloride increase **solubility** of both polar and nonpolar molecules in rough proportion of their accessible surface areas.

Temperature:

a) Most proteins can be denatured by heat, which affects the weak interactions in a protein (primarily **hydrogen bonds**) in a complex manner.
b) If the temperature is increased slowly, a protein's conformation generally remains intact until an abrupt loss of structure (and function) occurs over a narrow temperature range.
c) The abruptness of the process suggests that unfolding is a **cooperative** process: loss of structure in one part of the protein destabilizes other parts. The temperatures at which various proteins unfold vary enormously.

Note: The effects of heat on proteins are not yet readily predictable. The very heat-stable proteins of thermophilic bacteria have evolved to function at temperatures of hot springs\ (100°C). Yet the structure of these proteins differ only slightly from those of homologous proteins derived from bacteria such as *Escherichia coli*. How these small differences promote structural stability at high temperature is

not yet understood. Often a **single amino acid replacement** (**point mutation**) affects the stability of a protein by rendering it highly sensitive to changes in pH or temperature of its immediate environment.

PROTEIN AS A DYNAMIC MOLECULE: CONFORMATIONAL CHANGE

1) Protein molecule does not have a static rigid structure, instead its conformation can undergo large **coordinated changes** to move from one functional state to another.
2) Small differences in the environment such as **different pH** or the presence and absence of **ligands** can stabilize different conformational states of the protein.
3) These conformational changes can vary from adjustments of **side chain orientations** in the active site to **movements of loop regions**, differences in relative orientations of domains or changes in the quaternary structure of oligomeric proteins.
4) Such movements are usually **essential for function**, for enzyme catalysis, binding of antigens to antibodies, receptor-ligand interactions, muscle action and energy transduction and so on.
5) Some enzymes are known to undergo conformational changes on binding ligands or substrates.

Example:

Calmodulin

a) Calmodulin is a ubiquitous calcium-binding protein that is involved in a range of calcium-dependent **signaling pathways**.
b) Calmodulin binds to a variety of proteins such as **kinases**, **calcium pumps** and proteins involved in **motility**, thereby regulating their activities.
c) The calmodulin-binding regions of these proteins, comprising about 20 sequentially adjacent residues, vary in their amino acid sequences but they all have a strong propensity to form α helices.
d) Structure determination of calmodulin alone and of complexes with peptides have shown that peptide binding induces a large **conformational change** in the calmodulin molecule.
e) When calmodulin binds a ligand, only five groups actually change their conformation. They are five consecutive residues in an α **helix**, which unwind and turn into a **loop region**.
f) This rather small local change in peptide conformation causes one of the largest ligand-induced inter-domain motions known in a protein.

PROTEIN STRUCTURE I & II

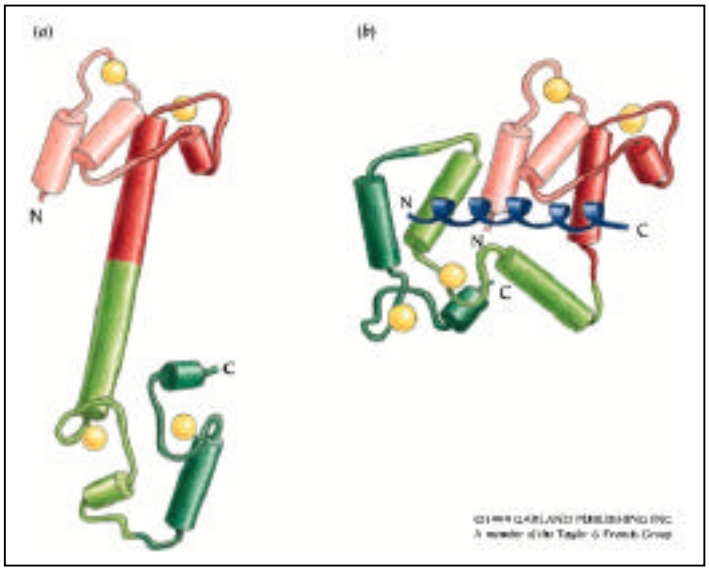


Figure 6.21 (Branden & Tooze). Schematic diagram of the conformational changes of calmodulin up on peptide binding. **(a)** In the free form the calmodulin molecule is dumbbell-shaped comprising two domains each having two EF hands with bound calcium. **(b)** In the form with bound peptides the helix linker has been broken, The two ends of the molecule are close together and they form a compact globular complex. The internal structure of each domain is essentially unchanged. The bound peptide binds as an α helix.

MEMBRANE PROTEINS

1) Biological membranes are lipid bilayers that are hydrophilic on the two outer surfaces and hydrophobic in between.
2) In membrane proteins, the **hydrophobic segments are embedded in the** membrane with the hydrophilic portions of the protein molecules projected on the either side of the bilayer.

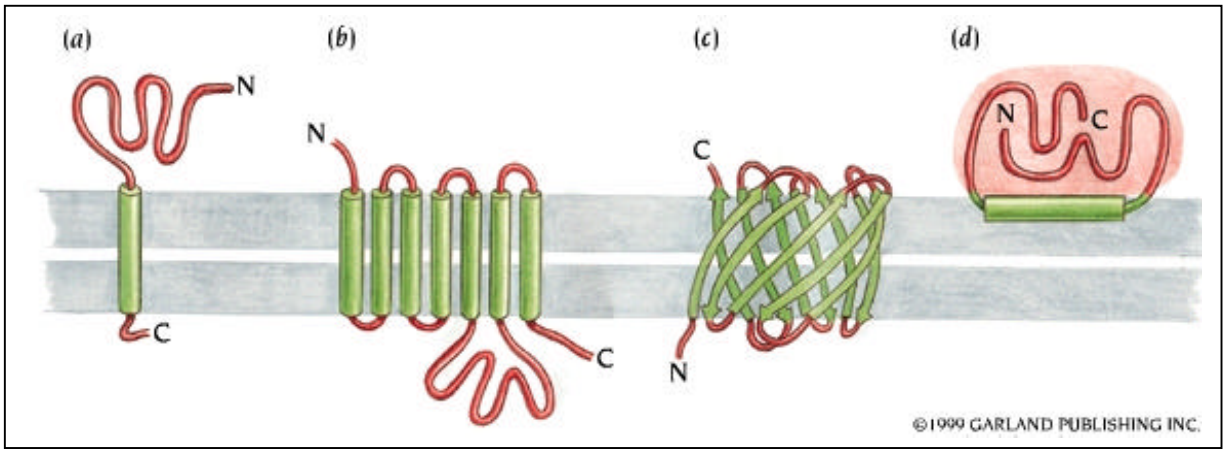


Figure 12.1 (Branden & Tooze). Four different ways by which protein molecules may bind to a membrane. Alpha-helices are drawn as cylinders and β strands as arrows. From left to right are **(a)** a protein whose polypeptide chain traverses the membrane once as an α helix, **(b)** a protein that forms several transmembrane α helices connected by hydrophilic loop regions, **(c)** a protein with several β strands that form a channel through the membrane, and **(d)** a protein that is anchored to the membrane by one α helix parallel to the plane of the membrane.

3) In some proteins polypeptide chain traverses the membrane only once (as in hemagglutinin and neuroaminidase of influenza virus).
4) The polypeptide chain of other membrane proteins passes through the membrane several times **usually as α helices** (as in the case of **Bacteriorhodopsin**) but in some cases as β **strands** (as in bacterial outer membrane protein **Porin**). Some proteins do not traverse the membrane at all, instead are attached to one side of the membrane either through α **helices** that lie parallel to the membrane surface or by **covalently liked fatty acids** to the protein that intercalate the lipid bilayer of the membrane.

Hydropathy plot

a) Based on experimental data and theoretical approaches, each amino acid side chain within a transmembrane α helix has different hydrophobicity (Apathy for polar solvent).
b) These **hydrophobicity scales** can be used to identify potential **transmembrane helices**.

Table 12.1 Hydrophobicity scales

Amino acid	Phe	Met	Ile	Leu	Val	Cys	Trp	Ala	Thr	Gly	Ser	Pro	Tyr	His	Gln	Asn	Glu	Lys	Asp	Arg
A	2.8	1.9	4.5	3.8	4.2	2.5	–0.9	1.8	–0.7	–0.4	–0.8	–1.6	–1.3	–3.2	–3.5	–3.5	–3.5	–3.9	–3.5	–4.5
B	3.7	3.4	3.1	2.8	2.6	2.0	1.9	1.6	1.2	1.0	0.6	–0.2	–0.7	–3.0	–4.1	–4.8	–8.2	–8.8	–9.2	–12.3

Row A is from J.Kyte and R.F. Doolittle; row B, from D.A. Engelman, T.A. Steitz, and A. Goldman.

c) For each position in the sequence, a **hydropathy index** can be calculated.
d) In transmembrane helices the **hydropathy index is high** for a number of consecutive positions in the sequence.
e) **Charged amino acids** are usually absent in the middle region of transmembrane helices because it would cost too much energy to have a charged residue in the hydrophobic lipid environment.
f) It might be possible, however, to have two residues of **opposite chare** close together in the lipid membrane because they neutralize each other.
g) When the hydropathy indices are plotted against residue numbers, the resulting curves, called **hydropathy plots**, identify possible transmembrane α helices as broad peaks with high **positive values**.

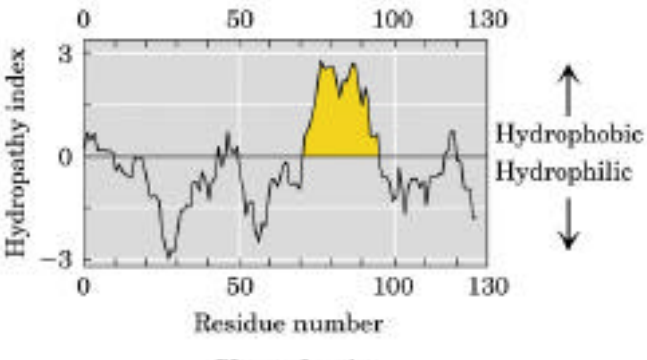


Figure 12-17 (Nelson and Cox). Hydropathy plots. Hydropathy index is plotted against residue number for two integral membrane proteins. (a) Glycophorin from human erythrocytes has a single hydrophobic sequence. (b) Bacteriorhodopsin, known from independent physical studies to have seven transmembrane helices, has seven hydrophobic regions.